DOCUMENT RESUME

ED 337 496 TM 017 338

TITLE	Student Assessment: A Review of Current Practices and Trends in the United States and Selected Countries.
INSTITUTION	New York State Education Dept., Albany. Office for Planning, Research, and Support Services.
PUB DATE	Jun 91
NOTE	84p.
PUB TYPE	Reports - Evaluative/Feasibility (142)
EDRS PRICE	MF01/PC04 Plus Postage.
DESCRIPTORS	*Educational Assessment; Educational Change; Educational Improvement; Educational Policy; Educational Practices; Elementary Secondary Education; *Evaluation Methods; Foreign Countries; Portfolios (Background Materials); *State Programs; *Student Evaluation; *Testing Programs; Test Use; Trend Analysis
IDENTIFIERS	Authentic Assessment; New York; *Performance Based Evaluation; United States

ABSTRACT

A comprehensive overview of the current state of student assessment and a glimpse of the extent of student assessment in New York State (NYS), throughout the nation, and in foreign countries are provided. Background on existing testing programs in NYS and the assessment components of "A New Compact for Learning," a set of proposals from the NYS Board of Regents for educational improvement, are included. Trends are identified, and significant reports or studies are reviewed. Test reform activists have initiated a campaign to limit the mass testing of young children in NYS, but in general, NYS's testing policies/practices agree with desirable testing practices. The development of large-scale performance-based assessment is now underway throughout the nation. Although there are some concerns that the new assessment is moving too fast, proponents argue that to determine its worth, performance-based testing must be initiated on a large scale. Proposals of the "New Compact for Learning" with regard to student portfolios and review of testing show a willingness to explore further change. A chart lists alternative assessment practices for each state. Seven appendices provide: descriptions and scheduling of NYS assessment programs; an outline of norm-referenced and criterion-referenced measurement; a table outlining questions used in the NYS Testing Program; lists of advocates for replacing standardized tests with alternative testing practices and advocates for a ban on standardized testing of young children; an outline of data sources for the 50-state testing comparison chart; and a list of districts taking part in the Urban District Leadership Consortium Project. A 74-item bibliography and additional information sources is included. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED337496

STUDENT ASSESSMENT:

A Review of Current Practices and Trends in the United States and Selected Countries

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

ROBERT M. TROMBLY

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM017338

The University of the State of New York
The State Education Department

BEST COPY AVAILABLE

TM

STUDENT ASSESSMENT: A REVIEW OF CURRENT PRACTICES AND TRENDS IN THE UNITED STATES AND SELECTED COUNTRIES

THE UNIVERSITY OF THE STATE OF NEW YORK
The State Education Department
Office for Planning, Research and Support Services
June 1991

THE UNIVERSITY OF THE STATE OF NEW YORK
Regents of The University

Martin C. Barell, Chancellor, B.A., I.A., LL.B	Muttontown
R. Carlos Carballada, Vice Chancellor, B.S.	Rochester
Willard A. Genrich, LL.B.	Buffalo
Emlyn I. Griffith, A.B., J.D.	Rome
Jorge L. Batista, B.A., J.D.	Bronx
Laura Bradley Chodos, B.A., M.A.	Vischer Ferry
Louise P. Matteoni, B.A., M.A., Ph.D.	Bayside
J. Edward Meyer, B.A., LL.B.	Chappaqua
Floyd S. Linton, A.B., M.A., M.P.A.	Miller Place
Mimi Levin Lieber, B.A., M.A.	Manhattan
Shirley C. Brown, B.A., M.A., Ph.D.	Albany
Norma Gluck, B.A., M.S.W.	Manhattan
Adelaide L. Sanford, B.A., M.A., P.D.	Hollis
Walter Cooper, B.A., Ph.D.	Rochester
Carl T. Hayden, A.B., J.D.	Elmira
Diane O'Neill McGivern, B.S.N, M.A., Ph.D.	Staten Island

President of The University and Commissioner of Education
Thomas Sobol

Executive Deputy Commissioner of Education
Thomas E. Sheldon

Coordinator, Planning, Research and Support Services
Julio Vidal

The State Education Department does not discriminate on the basis of age, color, religion, creed, disability, marital status, veteran status, national origin, race, gender or sexual orientation, in the educational programs and activities which it operates. Portions of this publication can be made available in a variety of formats, including braille, large print or audio tape, upon request. Inquiries concerning this policy of equal opportunity and affirmative action should be referred to the Department's Affirmative Action Officer, NYS Education Department, 89 Washington Avenue, Albany, NY 12234.

TABLE OF CONTENTS

INTRODUCTION 3

STUDENT ASSESSMENT IN NEW YORK 4

Current Practice 4
Time for Change 5

NATIONAL TRENDS IN TESTING 7

Overview 7
Beyond the Multiple Choice "Bubble" 8
A Different View of Multiple Choice Testing 9
Other Testing Issues 10

PERFORMANCE-BASED TESTING 12

Types of Alternatives 12
Benefits of Performance-Based Testing 12
Need for Caution 13
Performance Testing in Other States 14
Performance Testing in New York State's Assessment Programs 31

NATIONAL PERFORMANCE ASSESSMENT PROJECTS 33

National Testing Initiatives and Project Development 33
Urban District Leadership Consortium Projects and Partnerships 34
Council of Chief State School Officers' Project 36
CRESST at University of California at Los Angeles. 37
Office of Technology Assessment - U.S. Congress 38

TESTING AROUND THE WORLD 40

Research Availability 40
Canadian Provincial Assessment Activity 40
Literature Review Findings 42
International Assessment of Educational Progress 44

CONCLUSION 45

APPENDICES:

A. New York State Assessment Programs: Descriptions and Scheduling 47

B. Norm-Referenced and Criterion-Referenced Measurement 49

C. Types of Questions Contained in the New York State Testing Program 51

D. Advocates for Replacing Standardized Tests With Alternative Testing Practices 52

E. Advocates for Ban on Standardized Testing of Young Children 55

F. Data Sources for 50 State Testing Comparison Chart 56

G. Urban District Leadership Consortium Project Participating Districts 57

BIBLIOGRAPHY 59

OTHER INFORMATION SOURCES 63

"Pupils will be assessed for progress toward proficiency and mastery as well as for minimum competency. The emphasis will be on individual progress toward fixed standards of performance, not on comparative ranking of performance among those in the same age or grade....

The existing State testing program, except for Regents examinations, will be revised. Assessments will not be limited to multiple-choice, standardized tests, but will include components such as the following:

- *Examinations which measure problem-solving skills and the ability to analyze and synthesize, as well as factual recall.*
- *A portfolio of the pupil's best work, certified by his/her teachers and evaluated by qualified raters.*
- *A professional evaluation of the pupil's accomplishments, made by his/her teachers. This evaluation should extend not only to basic skills and knowledge, but also to desirable qualities (such as persistence, creativity, and sensitivity to others) not easily measured by conventional means....*

The revised State testing program will intrude as little as possible upon the time available for instruction. Where feasible, sampling techniques will be used. State assessments will be consolidated at grades 4, 8, and 12. Local assessments of programs will be made as appropriate throughout the grades...."

---New York State Board of Regents
"A New Compact For Learning"
March 1991

INTRODUCTION

Student testing is an integral component of the schooling process with the results used by teachers to measure pupil progress and to identify the student's strengths and weaknesses. Outside the school environment, policymakers, the business community and the public look beyond the individual student's scores by using test results to pass judgment on the schools, school districts, and the statewide condition of education. The emphasis on educational accountability during the 1980s precipitated the call for more extensive testing. Scrutiny of student assessment practices in America is currently recognized as an important piece in unraveling the puzzle of restructuring our education systems to meet the challenges of the 1990s.

Test reformers believe that certain tests (standardized or multiple choice format tests) in widespread use do not adequately serve the student, the teacher, or the school district since those tests measure only a small part of the curriculum and reveal only a portion of the student's abilities. With the "high stakes" decisions that now are coupled to test scores, such as remediation, "gatekeeping" and funding, there is concern that the practice of "teaching to the test" has hampered the development of higher order learning since assessment tools are known to "drive" curriculum and instruction. There has been a surge of interest in finding different measurement instruments to gain greater insight into student capabilities. The use of alternative assessments, often termed "performance-based assessment" and/or "authentic evaluation," has gained strong support.

This paper provides a comprehensive overview of the current state of student assessment while providing a glimpse of the extent K-12 students are being assessed in the State, throughout the nation and in foreign countries. Background on the existing testing programs in New York State and the assessment components of **A New Compact for Learning** are included to serve as a frame of reference for this review of the testing picture outside of New York State--in the other states and foreign countries. Trends are identified and significant reports or studies described.

The emerging significance of performance-based testing warrants the special attention given to an exploration of the extent of interest, development and use of such testing practices in the United States and abroad. In addition to terminology and examples, the pros and cons of this innovation are detailed. Finally, information is included on national projects relating to implementation of performance-based testing.

STUDENT ASSESSMENT IN NEW YORK

Current Practice

Since the initiation of the Regents eighth grade exam in 1865, New York State's students have been taking State tests as a means of assisting colleges with admission decisions. The high school Regents examination program followed in 1878. The Pupil Evaluation Program (or PEP test) was a response to requirements of the 1965 Federal law that provided funding for disadvantaged students and, beginning in 1979, PEP scores have been tied to State-funded remediation programs. High school level competency tests for graduation were first administered during the 1979-80 school year. Expansion of competency testing and new program evaluation and proficiency tests were components of the 1984 reform effort, the Regents Action Plan.

As of the 1990-91 school year, there are seven distinctive testing programs in place in New York State. The Regents Examinations, administered in grades 9-12, test students in college preparatory courses in English, mathematics, science, social studies and foreign languages. Successful completion of these examinations, and other requirements, leads to a Regents Diploma upon high school graduation. Students earning local diplomas must pass Regents Competency Tests in reading, writing, mathematics, science and social studies. In order to identify students in need of remediation, low-performing students must take the Preliminary Competency Tests in reading and writing in grades 8 or 9. Students in occupational courses take a variety of achievement tests (Occupational Education Proficiency Examinations) in vocational subjects during grades 9 through 12 to demonstrate competencies for Regents or local diploma requirements. Students may receive high school credit for their study of foreign language in elementary and middle school by successful performance on the Second Language Proficiency Examinations. All students in grades 3 and 6 take Pupil Evaluation Program (PEP) tests in mathematics and reading. Writing tests are administered in grade 5 to identify students in need of remediation, and for school assessment/accountability purposes. Program Evaluation Tests are given to all students in science (grade 4) and social studies (grades 6 and 8) to assess the quality of the school's programs. The current statewide assessment program outlined in Appendix A provides individual testing program descriptions and a chart depicting their scheduling by grade-level. A March 1989 New York State Education Department report, "New York State Testing Program for Elementary and Secondary Schools," identifies the following characteristics as shared by these State test programs:

- based on State-recommended or prescribed courses of study;
- intended to establish and maintain achievement standards; and
- intended to provide elementary and secondary school accountability.

Unlike many states, New York's State-level testing program contains no norm-referenced tests (NRTs); the tests in use are criterion-referenced (CRTs).[1] Briefly, CRTs reflect what the students know or how well they perform in absolute terms, while NRTs are given to a sample group of students whose grades are used as the norm for future scoring. Appendix B provides more information on the meaning of these terms and compares the qualities of each test type.

In New York State, teachers are integral to test development. Committees of teachers plan, write questions, revise and assemble, as well as review the final versions of each test. The 1989 SED testing report noted that Department-developed tests are not the practice in most states. Just one test offered to NYS students, the Degrees of Reading Powers Test, is not SED-developed. In addition to the State-administered examinations, some local school districts administer norm-referenced standardized achievement, ability, Intelligence Quotient (IQ) and aptitude tests to make decisions on grade promotion and to track by ability.

Time For Change

Approved by the New York State Board of Regents, **A New Compact for Learning** is a set of proposals to improve educational results for the State's elementary, middle, and secondary students. It offers specific suggestions for changes to statewide assessment practices. These proposals respond to the Compact's principles, emphasizing results and content mastery.

The Compact calls for assessment based on both minimum competency and evaluation for mastery levels. Specifically, the Compact proposes that performance standards for both minimum competency and mastery be established for each subject in grades 4, 8, and 12 with all students encouraged to excel. There is a focus on individual progress toward an established standard of performance for those students in the same grade or of the same age. The Compact establishes that while the Regents Examinations will continue, other existing testing programs may be replaced or revised. The proposal calls for each pupil in these grades to be evaluated by a three-part assessment process that would include portfolio work, teacher evaluation of the student's desirable qualities, and core subject examinations to measure problem-solving skills and factual recall. The Compact notes that new examinations should not be limited to multiple choice, standardized tests.

---

[1]While SED uses the Degrees of Reading Power test which does have norms, the norms are not used by the Department.

Also under consideration is a proposal that high school graduation should be tied to the satisfactory completion of a secondary school assessment and the prescribed number of courses. This new assessment is envisioned as a series of measures to be taken until the established standard is met.

The Compact's assessment proposal reflects certain of the current national/international trends in testing. The following review of states' actions and international assessment information demonstrates that examination of new alternatives and a review of New York State's existing assessment programs are consistent with the ferment in testing seen across the country and beyond.

NATIONAL TRENDS IN TESTING

Overview

Standardized testing has dominated the testing scene since its introduction in the 1940s. Newsweek, in a September 1990 special edition, **Education: A Consumer's Handbook**, estimated that 127 million tests are given each year, with norm-referenced tests identified as the most widely used. These tests are easily administered, machine-scored and not expensive.

With the waves of education reform in the 1980s came an increased demand for accountability. Typically, more testing was the response, and nationally there was an increase in such standardized tests. A certain disillusionment about standardized tests followed the appearance in 1987 of what is now known as the Lake Wobegone effect--the aftermath of J.J. Cannell's study, **Nationally Normed Elementary Achievement Testing in America's Public Schools: How All Fifty States Are Above the National Average**. The public began to question how all states could be above-average.

Cannell credited the Associated Press with coupling the report to Lake Wobegone, "Garrison Keillor's mythical town in Minnesota where 'all the women are strong, all the men are good looking, and all the children are above average.'" The initial study concluded that such testing had in fact resulted in 70 percent of U.S. elementary students, 90 percent of the nation's school districts, and all 50 states testing above the national norm. In effect, such "feel good" tests lead to an illusion of improved achievement even among the most marginal school districts in the country. A 1989 follow-up report, **How Public Educators Cheat on Standardized Achievement Tests**, detailed the flawed use of "Wobegone" tests in achieving these results. The overstatement of achievement is a product of numerous conditions including the use of publishers' tests with old or outdated norms; the pressure on schools and teachers to cheat generated by high stakes testing for accountability purposes; lack of appropriate test security measures; misleading reporting methods; and use of deceptive test practices such as excluding special education and/or bilingual students from such testing.

Recently, student assessment reform activities center on three issues: the surge of interest in alternative testing practices; calls to limit the testing of young students; and, in some states, concern that students are overtested.[2]

---

[2]In New York State's proposed "A New Compact For Learning," the assessment provisions call for new individual-oriented measurement practices, and a reevaluation of present practices in the statewide testing program.

Beyond the Multiple Choice "Bubble"

There has been a major campaign by test reformers to move away from reliance on standardized, multiple choice (both NRT and CRT) testing in favor of alternative testing practices. The move toward alternative testing is the assessment issue most closely related to the Compact.

As noted previously, New York's statewide-level testing program is not norm-referenced. However, the existing New York State program does use multiple choice for many of the tests. (Appendix C provides information on the extent of multiple choice items used in New York State's assessment program.) Proponents of test reform are lobbying for the use of imaginative and, many feel, fairer measures of achievement. (See "A Different View of Multiple Choice Testing" for the other side of this issue.) In response, the Compact is advancing testing proposals that do not rely exclusively on multiple choice format, i.e., portfolio use and teacher evaluations.

The National Center for Fair and Open Testing (FairTest), a well-organized and vocal public watchdog group, has campaigned nationally to reduce the use of standardized, multiple choice testing. The Center has lobbied for the development and use of alternative testing mechanisms utilizing other than the multiple choice/short answer format. FairTest has gained the support of a number of important national groups and individual educators who have endorsed their "Statement on Genuine Accountability." Originally, this action to reduce standardized test use was organized to dissuade the President and governors from overreliance on standardized tests as a way to measure progress toward Education Summit Goals in September 1989. Among those endorsing the statement were the National School Boards Association, both the United Federation of Teachers and New York City's American Federation of Teachers, Bank Street College of Education, the National Parent-Teachers Association, and the National Association for the Advancement of Colored People. (Appendix D provides the complete list of endorsers.)

Many reports and journal articles concur that: standardized testing is overdone and misused as an accountability mechanism; and, education reform should focus on assessment based on what the student can do rather than student scores on such multiple choice tests. These reports include those from Archbald and Newmann (1988), Harris and Sammons (1989), Medina and Neill (1988), and FairTest (1990). Support was found in journal articles by Haney and Madaus (1989), Shepard (1989), and Wiggins (1989).

The most recent such report, titled **From Gatekeeper to Gateway: Transforming Testing in America**, was issued after three years of study by the National Commission on Testing and Public Policy. Funded by the Ford Foundation, the report examines the weaknesses inherent in the use of standardized test scores for minority test-takers. Study

findings indicate that standardized testing requirements may unfairly inhibit capable minorities from realizing their full potential.

Gateway urges the redirection of testing from multiple choice tests, and expansion of research and development programs in the new types of assessments that recognize the talents of all people. "Important decisions about people and institutions should, where feasible, be based on multiple sources of information, especially direct evidence of actual performance in school or on the job. Thus candidates should supply answers, perform acts, demonstrate skills, create products, and supply portfolios. Previous accomplishments should also be considered." (p. x)

A Different View of Multiple Choice Testing

Attention has focused mainly on the trend away from the multiple choice format for testing. However, not all testing practictioners support this move. Feinberg, writing for the College Board's **Commentaries** in 1990, notes that standardized tests were once considered "alternatives" to the traditional essay exam which was criticized for use of arbitrary topics and inconsistent grading practices that were influenced by such "irrelevances" as neat handwriting. Denying that standardized tests are biased against race, sex and social class, Feinberg asserts that the test format is not the problem with multiple choice tests. Charges that such tests are too simple ignore the fact that the tests can be analytical. The problem lies with the choice of the test; with the politics of testing leading to the selection of easier tests by schools for "obviously self-serving reasons."

The practices that test reformers criticize are a result of the manner that the multiple choice tests are "normed, administered and used, not an inevitable consequence of the format." Accountability concerns have resulted in standardized tests now being used to evaluate schools and school systems rather than just the student. Feinberg recommends that for these purposes, a test which samples students and uses mostly multiple choice questions from a large test bank (similar to the National Assessment of Educational Progress (NAEP)) is appropriate. Such a test would make cheating difficult and yet test "broad swatches of curriculum."

The main advantage of the multiple choice format is that such tests are able to test a great range of topics in a brief period of time. Feinberg acknowledges the other problems associated with the multiple choice format and suggests that concerns can be addressed by:

- regularly giving new tests;
- creating new norms based on actual performance; and
- imposing tighter test security.

While detailing how multiple choice testing can become better, the Feinberg article also addresses the problems associated with the new alternative tests. These problems are among those identified in a following section, "Caution Needed."

In conclusion, Feinberg notes that "Although the denunciations are stern and the controversy sharp, there may soon be a place in large-scale testing for both multiple-choice and some of the performance alternatives [though few of the alternatives are actually ready for use]. But despite fervent claims, it seems unlikely that any change in test format will have as much impact as harder test content and higher test standards might have."

Other Testing Issues

Also of consequence in assessment trends is the call to reduce testing. Two watchdog groups are trying to curb the testing of young children in grades Pre-K through 2, and to examine the total testing picture in some states to determine if over-testing is taking place.

The 1990 **Gateway** study reports on the prevalence of testing of young children:

> The Commission finds that testing primary school children for entry to or exit from a grade is poor education practice. Nonetheless, prekindergarten tests are mandated in more than 16 states, widely used in seven states, and known to be used at the district level in more than 37 states. Kindergarten exit/first grade entrance tests are used in at least five states and known to exist at the district level in an additional 37. In some school districts as many as 60 percent of the kindergartners are judged to be "unready" for first grade because of their scores on a 'readiness' test. Achievement testing is required for first graders in nine states; for second graders in nine states; and for third graders in 27 states. (pp. 14-15)

FairTest, in conjunction with the New York Public Interest Research Group (NYPIRG), has initiated a campaign in New York State requesting the Board of Regents to act against local school district use of standardized testing for young children below age 8. The State's present testing program does not test this age group, yet FairTest and NYPIRG contend that many local school districts do have such practices. Their 1989 report, **Failing Our Children: How Standardized Tests Damage New York's Youngest Children,** surveyed the use of such tests in the State's school districts. Based on a 20 percent response rate, the study concluded that the percentage use of standardized tests by grade level is: Kindergarten -- 75 percent, Grade 1 -- 88 percent and Grade 2 -- 93

percent.[3] A Fall 1990 follow-up report, **Standardized Tests and Our Children: A Guide to Testing Reform in New York**, urges a ban on testing of young children and promotes the use of authentic (performance-based) evaluation. In October 1990, the FairTest/NYPIRG coalition sent an appeal to the Board of Regents and the Commissioner of Education requesting a ban on mass standardized testing for young children below age 8 by the fall of 1996, as well as funding for the implementation of developmentally appropriate assessments and teacher training. (Endorsers of this initiative are listed in Appendix E.) Such states as Georgia, North Carolina, Mississippi, Arizona and Texas have adjusted their testing policies by changing the type of test or eliminating mandated testing for this age group.

The **Gateway** study provides evidence of the scope of the testing explosion: "Mandatory testing includes some 20 million school days and the equivalent of $700 to $900 million in direct and indirect expenditures annually." (p. x) These 20 million days relate only to elementary and secondary level schooling and count only the time used in actually taking the standardized tests. The count does not include the considerable additional time spent in classrooms on test preparation. The National Commission on Testing and Public Policy, author of the study, has stated that this is a conservative estimate. The 127 million standardized tests previously reported by Newsweek are the same amount cited in this report.

The overtesting of students seems to have resulted from the emphasis that continues to be placed on accountability. Chris Pipho of the Education Commission of the States (ECS) finds some states reconsidering the amount of testing that is in place and raising concerns about the amount of time devoted to testing. Here again, there is evidence that as some states' testing programs have been enlarged to respond to accountability concerns by policymakers and the public, there has been no reduction in the level of testing of the standardized type at the local level. When Federal testing requirements, such as Chapter I, are added to those of local and state authorities, the result is multiple layers of testing. In 1989, Illinois passed legislation limiting statewide student testing time to 5 hours per year; Texas is looking into this issue.

---

[3]Critics note that the low response rate is problematic since it generalizes to the population based upon inadequate sample responses and is therefore flawed.

PERFORMANCE-BASED TESTING

Types of Alternatives

The alternatives to standardized testing are referred to by the following names: authentic evaluation, appropriate evaluation, performance-based evaluation and direct assessment. In general, these terms are used to describe many alternative testing strategies including the use of portfolios, essay questions, performances, tasks, exhibitions, open-ended test items, experiments, teacher observations, checklists and inventories. Each of these measurement tools is intended to demonstrate what the student can do rather than how they score on an examination. The most prevalent term in use is "performance-based testing."[4]

Benefits of Performance-Based Testing

Testing reform advocates are enthusiastic about the need to gather more information on the alternative tests and are actively pushing for the inclusion of new varieties of tests as a significant component of the total reform/restructuring of education. In addition to FairTest, the testing experts based in centers for study such as the National Commission on Testing and Public Policy (Boston College), Research for Better Schools (Philadelphia), Council for Basic Education (Washington, D.C.), Center for Research, Evaluation, Standards and Student Testing (University of California at Los Angeles), and the Urban District Leadership Consortium (Washington, D.C.) are among the numerous groups that are actively promoting and studying performance-based testing. In order to collect information relating to the study of such new assessment practices, test reformers believe it is imperative that specific initiatives are used on a large scale in order to provide the necessary data for evaluation purposes.

Among the positive attributes of performance-based testing, proponents assert that this form of testing:

- leads to improved curriculum and teaching since its use will spur instruction in higher order skills;
- eliminates the narrowing of the curriculum found with standardized test use;
- responds positively to "teaching to the test" since teaching to the performance-based test will result in better instruction and curriculum;

---

[4]The initiatives on portfolio and teacher evaluations in "A New Compact For Learning" are examples of performance-based testing.

- results in a greater understanding of the student's abilities than is provided by the standardized type of test;
- alleviates the boredom of classroom instruction keyed to standardized test driven instruction and replaces "multiple choice teaching" with a "thinking curriculum";
- provides teachers with professional growth and gives them the opportunity to see the effects of their instruction in the performances of their students;
- contributes to advancement of teacher empowerment (greater decision-making authority) by expanding the participation of teachers in the development of performance assessment programs and by providing teachers with an active role in the scoring process for such alternative assessments; and
- holds promise for use in the restructuring process of urban center education systems since alternative testing results in a greater understanding of the students' abilities than is provided by the standardized test.

Need For Caution

The ground swell of support for a shift to these new types of assessments has resulted in concerns, cautions and reservations being expressed by those who think that things are moving too fast in this new direction. Some of the cautions are raised by individual advocates for performance-based testing who are alarmed that the new brand of testing is being expected to do too much. Those who have raised questions include some academic psychometric experts, such as Professor Edward Haertel of Stanford University; the National Science Academy's panel on international comparative education; Christopher T. Cross, Assistant U.S. Secretary of Educational Research and Improvement; Chester Finn of Vanderbilt University (former research head at the U.S. Department of Education); H.D. Hoover, Director of the Iowa Basic Skills Test; and parents.

The cautions put forth include that these new alternative/performance assessments are:

- untried and may not prove to be a better measurement of student ability;
- considerably higher in cost than standardized testing;
- identified by some psychometric experts as less reliable and less valid than traditional tests;[5]

---

[5]**Standards For Educational and Psychological Testing**, the source of definitions used by testing professionals, explains "reliability" as "The degree to which test scores are consistent, dependable, or repeatable, that is, the degree to which they are free of error."

- being asked to do too much since, in addition to measuring student abilities, the new tests will be used as an accountability system for individual schools and the school district;
- just as vulnerable to being "taught to" as multiple choice style testing;
- liable not to result in the better performance by those who fare poorly on the standardized tests since performance testing reflects reality;
- likely to reveal wide gaps in the achievement of disadvantaged students since changing the type of assessment will not change outcomes;
- unable alone to improve instruction but must be part of a comprehensive strategy that raises standards on what pupils should be able to accomplish;
- necessitate that teachers will be required to teach in new ways consequently requiring professional development and reeducation;
- not useful to policymakers since they will be able to draw few conclusions about students' overall performance because only a relatively few number of tasks are used in performance evaluation;
- the cause of narrowing curricula in cases where the performance test may be restrictive, e.g., the production of a social studies docu-drama relating to a specific limited historical time period;
- cause for concern for some parents who have expressed alarm that performance testing may hamper their children's admission to college;
- thought to be too subjective by minority parents worried that biases might surface;
- subject to lawsuits which may result when performance tests are used to determine student advancement or rewards such as for scholarship purposes or the withholding of drivers' licenses; and
- more vulnerable to bias charges because performance assessments use fewer test items than standardized multiple choice tests and are therefore unable to "balance out" the unfairness of testing.

Performance Testing In Other States

Many states have shown interest in alternative testing. Two surveys have been completed that examine the extent of performance-based testing activities in the states. The first such survey was made in the spring of 1990 by Pamela Aschbacher of UCLA's Center for Research, Evaluation, Standards, and Student Testing (CRESST). A more extensive type survey and analysis conducted by the Council of Chief State School Officers

---

The same source defines "validity" as "The degree to which a certain inference from a test is appropriate or meaningful."

(CCSSO) was completed in December 1990 using a broader definition of performance testing than the CRESST survey.

The CRESST survey, made public at the 20th Annual Assessment conference sponsored by the Education Commission of the States and the Colorado Department of Education in June 1990, reported that the most prevalent form of performance-based testing was the direct writing assessment. States were asked to identify those testing activities that are "alternatives to multiple choice testing but [this definition] does not include writing assessments unless portfolios, integrative, or other innovative approach [are] used." Based on this criterion, Aschbacher identified the following states as having some form of performance testing in place: Delaware, Hawaii, Maine, Massachusetts, Michigan, New York and North Carolina. CRESST listed an additional 16 states that were either planning or exploring the use of performance testing and 27 states as having no plans.

The December 1990 CCSSO survey findings reported wider use of performance-based testing in part due to the broader definitions used for this survey. In fact, an accompanying CCSSO document, "Models of Performance Assessment," states that "In practice, performance assessment is anything NOT multiple-choice or machine scorable." The survey document provided the following more specific definitions. Performance assessment was defined as "ways of assessing student and program achievement which require direct demonstration of the target knowledge and skill...includ[ing] direct writing samples; open-ended questions; demonstrations; experiments; and group projects." The other identified types of performance assessments identified in the CCSSO matrix included portfolio (reviewed selections of student work accumulated in a folder over a period of time) and "enhanced multiple choice" (multiple choice items which are not simple recognition of a correct answer, but require analyzing a problem before selecting an answer). Using these definitions, the findings are significantly different from the earlier CRESST survey. CCSSO reports that 33 states currently have some types of performance-based testing activities implemented. An additional five states have pilot or field test activities under way and two other states are in the planning stage with some details yet to be decided. Of the 33 states with existing programs, 13 have additional performance activities in the development or planning stages. Just 10 states were listed as having no performance assessment activity.

The accompanying State Education Department-developed chart moves beyond the CRESST and CCSSO survey by providing--along with information on performance-based testing--additional student assessment information to present a more complete picture of what activity is taking place or planned in all 50 states. Using several sources, including the previously described surveys (see Appendix F for source identification), this chart provides a comprehensive view of the extent of testing carried on by each state, the types of testing (listed under "traditional" when in common use), and when applicable, activity relating to performance-based assessments. In those states where specific information was

available, the nature of the test (CRT or NRT) has been included. The following findings are based on data detailed on the Chart on Alternative and Traditional Assessment Practices in the United States:

- 11 states test pupils below the grade 3 level using traditional testing; 6 states test this age group via performance assessment; 3 states use both types of assessment practices and 3 states are developing alternative assessments;
- 20 states use exit exams, most frequently tied to high school graduation;
- 15 states test for minimum competencies; and
- 34 states have some form of performance-based assessment activity in place.

Of the state entries that distinguish the nature of the tests in use:

- 32 states utilize NRT tests;
- 23 states use CRT tests; and
- 15 states use both NRT and CRT tests.

Of the planned activities listed:

- 2 states plan to add NRT testing;
- 4 states plan to add CRT testing;
- 25 of the 34 states that have some form of performance-based assessment practices are planning/developing or discussing expansion of alternative assessment;
- 9 of the 16 states without existing performance-based assessment practices are planning/developing programs, and 2 other states are discussing alternative assessment; and
- 5 states have expressed no interest in moving toward alternative assessment activity.

CHART ON ALTERNATIVE AND TRADITIONAL ASSESSMENT PRACTICES IN THE UNITED STATES

STATE	ALTERNATIVE ASSESSMENT PRACTICES	TRADITIONAL ASSESSMENT PRACTICES						
	PERFORMANCE-BASED ACTIVITIES	CRT	NRT	Min Comp	Exit Exam	Level	Grade Level	Test Subject Areas/Comments
AL	◊ For 1991-92, the 3-year phase-in of an accreditation system that includes performance testing with portfolio use ◊ Writing assessment (grade 5) ◊ Integrated reading and writing format test and integrated math and science performance assessment (grade 2) ◊ Math performance assessment (grades 8-12)			*			3,6,9	Reading, math, language
					*	HS		Reading, math, language
			*				4,8	Reading, math, social studies, science
AK	* Pilot writing assessment (grade 10) * Pilot portfolio assessment projects ◊ Expansion of writing assessment to other grades		*				4,6,8	Reading, math, language arts
AZ	◊ Performance assessment for reading, writing, math (grades 3, 8, 12) to be fully in place for fall 1991 ◊ Portfolios		*				2-11	Reading, math, grammar
AR	# Adding writing assessment to existing tests (grades 7, 10)	*		*			3	Reading, math
		*			*	8	6,8	Reading, math, language arts, social studies, science
			*				4,7,10	Achievement
		#	#					Use of combined format

* In Place
◊ Planned/Under Development
\# Under Discussion

CRT Criterion-Referenced Test
NRT Norm-Referenced Test
Min Comp Minimum Competency Test

STATE	ALTERNATIVE ASSESSMENT PRACTICES	TRADITIONAL ASSESSMENT PRACTICES						
	PERFORMANCE-BASED ACTIVITIES	CRT	NRT	Min Comp	Exit Exam	Level	Grade Level	Test Subject Areas/Comments
CA	California Assessment Program: * Direct writing assessment (grades 8, 12) * Open-ended items are part of math assessment (grade 12) * Performance testing in physical education (grades 5, 7, 9) * Enhanced multiple choice use in math (grades 6,8,12) ◊ Performance assessments in science (grade 6) and social studies (grades 6,8,12) ◊ Integrated reading/writing performance assessments (grades 6,8) ◊ Portfolios for math (grades 1-12) and writing (grade 10) * Golden State Exams in history and economics include an essay and science exams include open-ended items and a performance task						3, 6, 8, 12 ◊ 10	CAP Tests: Reading, math, written expression
							8 ◊ 6, 12	Science, social studies
							8, 12 ◊ 3	Health
							9-12	Golden State Exams: end of course (voluntary)
							Local Administration Once: 4-6, 7-9, 10-11	Proficiency tests in reading, writing, computation
CO	* Student writing assessment (grades 4, 7, 10) ◊ Math/science problem-solving and/or performance measure		*				Local Choice	
CT	* Common Core of Learning (CCL) assessment (grades 9-12) piloted math and science performance assessments including portfolios, projects, speaking and listening activities * Direct writing and listening assessment are part of the Connecticut Mastery Test Program (grades 4,6,8) ◊ New law expands Mastery Tests to grade 10 and asks for new science and social studies components	*					All 4, 6, 8 Selective 9-12 ◊ 10	Mastery Tests: Reading, math, language arts, writing
DE	* Physical education and geography ◊ Writing assessment to begin in spring 1991 (grade 10) ◊ Interdisciplinary alternative assessments in early childhood assessment (grades 1 - 3)		* ◊ ◊				1-8, 11 ◊ 3 ◊ 6,8, 11	Reading, math, basic skills Reading, math, basic skills (spring 1991) Reading, math, basic skills, sciences social studies

* In Place
◊ Planned/Under Development
Under Discussion

CRT Criterion-Referenced Test
NRT Norm-Referenced Test
Min Comp Minimum Competency Test

STATE	ALTERNATIVE ASSESSMENT PRACTICES	TRADITIONAL ASSESSMENT PRACTICES						
	PERFORMANCE-BASED ACTIVITIES	CRT	NRT	Min Comp	Exit Exam	Level	Grade Level	Test Subject Areas/Comments
FL	* End-of-course exams are essay tests			*			3, 5, 8, 10	Reading, math, writing
	◊ New law initiates writing assessment component for each test and the opportunity to complete an optional career survey at grades 7, 10				*	HS		Communication skills, math
	# Science performance testing and portfolios	*					9-12	End-of-course exams
GA	* Georgia Kindergarten Assessment Program	*			*	3	1, 3, 6, 8	Reading, math
	* Writing assessment (grades 6, 8,10)						OPT	
	◊ New law drops grade 1 entrance test						2,4	
	◊ Voluntary science performance-based assessments (grades 1-4, 6)	*		*	*	HS	10-12	Reading, math, writing
			*				2, 4, 7, 9	Reading, math, language skills, social studies, science
HI	* Graduation exams have "hands-on" versions in reading, math, writing, social studies, language arts and vocational studies (grades 9 - 12)		*				3, 6, 8, 10	Achievement
		*			*	HS	9-12	Failing student must take performance version
	◊ Enhanced multiple choice and performance assessments in reading, math, science and social studies (grades 3,6,8,10)							
	# Art and physical education performance assessment							
ID	* Direct writing assessment (grades 8,11)		*		*	11	6, 8, 11	Local admin. requirements or score "C" for 14 core courses

* In Place
◊ Planned/Under Development
\# Under Discussion

CRT Criterion-Referenced Test
NRT Norm-Referenced Test
Min Comp Minimum Competency Test

STATE	ALTERNATIVE ASSESSMENT PRACTICES	TRADITIONAL ASSESSMENT PRACTICES						
	PERFORMANCE-BASED ACTIVITIES	CRT	NRT	Min Comp	Exit Exam	Level	Grade Level	Test Subject Areas/Comments
IL	* Writing samples (grades 3, 6, 8)	*					3, 6, 8,	Reading, math, language arts
	◊ Writing sample beginning in 1992 (grade 11)	*					11	Reading
	# Math, science and social studies performance assessment	◊					11	◊ Language arts (1992)
		◊						Science, social studies, fine arts, health & physical development (to be biennial)
IN	* Direct writing assessment (grades 3, 6, 8, 9, 11)	*					3, 6, 8, 11	Social studies, science
	◊ Limited tryout of performance testing in science, math (grade 6) and social studies		*				1, 2, 3, 6, 8, 9, 11	Language arts, math
IA	* Voluntary writing and listening assessment programs (grades 3-8) ◊ Portfolio assessment for local use	No Statewide Testing						
KS	◊ Enhanced multiple choice and performance math exams (grades 3,7,10) # Math portfolios	◊						Math
KY	◊ New reform legislation includes a performance-based assessment process by 1994-95 at the latest; also requires new interim program (see entry under traditional practices)		*				K, 1, 2, 3, 5, 7, 10	Reading, math, language, spelling, learning skills
			◊				◊ 4, 8, 12	◊ Reading, math, science, social studies, ("NAEP-Like", ready for 1991-92)

* In Place
◊ Planned/Under Development
\# Under Discussion

CRT Criterion-Referenced Test
NRT Norm-Referenced Test
Min Comp Minimum Competency Test

STATE	ALTERNATIVE ASSESSMENT PRACTICES	TRADITIONAL ASSESSMENT PRACTICES						
	PERFORMANCE-BASED ACTIVITIES	CRT	NRT	Min Comp	Exit Exam	Level	Grade Level	Test Subject Areas/Comments
LA	* Writing assessments (grades 5, 7, 10)						K	Dev. readiness
			*				4, 6, 9	Reading, math, language arts, study skills, social studies, science
		*					3	Math, language arts
							5, 7	Math, language arts, written composition
		*			*	HS	10	Math, English lang. arts, written comp.,
							11	social studies, science
ME	* Existing assessment program contains performance tasks for all academic subjects (grades 4,8,11) ◊ Voluntary integrated reading/writing (grades 8,11)						4, 8, 11	Reading, math, writing, social studies, science, humanities
MD	* Direct writing assessment is part of the graduation exam		*				3, 5, 8	Reading comp., math, lang. arts
	◊ Performance testing under development by a test publisher consulting with Maryland teacher representatives for reading, writing, math, science and social studies (grades 3,5,8,11)	*			*	HS	9-12	Reading, math, writing, citizenship
MA	* Direct writing assessment (grades 3, 6, 9)	*		*			3, 6, 9	Reading, math, writing
	* Math, science, social studies and reading using open-ended questions * Biennial performance testing in math and science (grades 4, 8)						4, 8, 12	Reading, math, social studies, science
	◊ Increased use of performance testing and open-ended questions							

* In Place
◊ Planned/Under Development
\# Under Discussion

CRT Criterion-Referenced Test
NRT Norm-Referenced Test
Min Comp Minimum Competency Test

STATE	ALTERNATIVE ASSESSMENT PRACTICES	TRADITIONAL ASSESSMENT PRACTICES						
	PERFORMANCE-BASED ACTIVITIES	CRT	NRT	Min Comp	Exit Exam	Level	Grade Level	Test Subject Areas
MI	* Performance testing in music, art, physical education; past activity in science and career development * Reading (grades 4, 7, 10) ◊ Performance assessment exams in math (grades 4, 7, 10), writing (grades 4, 5, 7, 8-11), science and social studies (grades 5, 8, 11), and employability skills (grades 8-12) for administration once every 3-4 years	*		*			4, 7, 10 5, 8, 11	Reading, math Science
MN	* Performance and enhanced multiple choice assessment in reading, math, writing and language arts (grades 6,9,11) ◊ Developing/planning performance assessments in science, social studies and science-3 day writing project							
MS	* Direct writing assessment (grades 8, 11)				◊	11	* 3, 5, 8, 11	Diagnostic tests (grade 11 to be used as exit exam)
MO	* Writing assessments field-tested and ready for spring 1991 implementation (grade 8) * Performance testing in all subject areas (grades K, 1)	*					2-10 (local choice)	Reading, math, language arts, social studies/ civics, science
MT			*				3, 8, 11	Reading, math, language arts, social studies, science (Local choice of test)
NB	◊ While there is no statewide testing program, the SEA is working with teachers on writing and math portfolios and art assessment		*				Once 4-6, 7-9, 10-12	Local decision
		*					At least by 5 (Local choice)	Reading, math, writing

* In Place
◊ Planned/Under Development
\# Under Discussion

CRT Criterion-Referenced Test
NRT Norm-Referenced Test
Min Comp Minimum Competency Test

STATE	ALTERNATIVE ASSESSMENT PRACTICES	TRADITIONAL ASSESSMENT PRACTICES						
	PERFORMANCE-BASED ACTIVITIES	CRT	NRT	Min Comp	Exit Exam	Level	Grade Level	Test Subject Areas/Comments
NV	* Writing assessments (grades 9, 11)		*				3, 6	Reading, math, language
							9	Reading, math
		*		*	*	HS	11	Reading, math, writing
NH					Local option		4, 8, 12	Math, communications
NJ	* Essay writing is a component of the graduation exam	*			*	HS	9	Reading, math, writing
	◊ Reading, math, writing (grade 8)				◊	HS	11	
					◊		8	Reading, math, writing (March 1991)
			*				3 - 8, 10,11	Reading, math, language arts (Chosen from State approved test list)
								◊Core subject areas: math (1992) science (1993) English (1994) social studies (1995)
NM	* Direct writing asessment (grades 4, 6)						1, 2	Reading
	◊ Voluntary enhanced multiple choice assessments in reading (grade 8), writing (grades 10-12), science (grades 3, 5, 8, 10-12), and language arts (grades 10-12) to be given at 4 year intervals		*				3, 5, 8	Reading, math, language, science, social studies, spelling, ref. skills
		*					3, 5, 8	Math, language arts, science, social studies
				*	*	HS		

* In Place
◊ Planned/Under Development
\# Under Discussion

CRT Criterion-Referenced Test
NRT Norm-Referenced Test
Min Comp Minimum Competency Test

STATE	ALTERNATIVE ASSESSMENT PRACTICES	TRADITIONAL ASSESSMENT PRACTICES						
	PERFORMANCE-BASED ACTIVITIES	CRT	NRT	Min Comp	Exit Exam	Level	Grade Level	Test Subject Areas/Comments
NY	* Direct writing assessment (grades 5, 8/9, 11) * Open-ended items are part of math assessment (grade 9)	*					3,6 5	Reading, math Writing
	* Direct reading assessment that measures inferential comprehension through the use of modified close technique (grades 3, 6, 8/9, 11)	*					4 6/8	Science Social Studies
	* Peformance testing in physical education (grades 4-12)	*					8/9	Reading, writing
	* Performance assessment is part of science assessment (grade 4) * Essays included in social studies tests (grades 6, 8, 10, 11) * Participation projects are part of social studies program evaluation (grades 6, 8) * Performance assessment (speaking, listening, and writing) is part of foreign language tests (grade 8)	*		*	*	HS	11-12 9 - 12 9 10 11	Reading, writing Math Science Global studies US history and government
	* Regents Examinations (RE): English RE includes listening and writing an essay based on literature and a composition based on a given situation or open-ended topic Modern Language RE include speaking, listening, and writing performance assessment	*		*	*	HS	9-12	Intro to Occupation, agriculture, business, health occupations, home economics, technology (for occ ed. sequence)
	Latin RE includes oral reading tasks						8	Second language proficiency (voluntary)
	Sequential Mathematics, Courses I, II, III include open-ended items Biology, physics RE include open-ended items Earth Science RE includes performance assessment	*			*	HS	9-12	16 subject end-of-course exams leading to Regents diploma (voluntary)

* In Place
◊ Planned/Under Development
\# Under Discussion

CRT Criterion-Referenced Test
NRT Norm-Referenced Test
Min Comp Minimum Competency Test

STATE	ALTERNATIVE ASSESSMENT PRACTICES	TRADITIONAL ASSESSMENT PRACTICES						
	PERFORMANCE-BASED ACTIVITIES	CRT	NRT	Min Comp	Exit Exam	Level	Grade Level	Test Subject Areas/Comments
NY	Prerequisite for taking any of four science RE is the participation in laboratory activities for at least thirty 40-minute periods and the submission of satisfactory written reports of these laboratory experiences Global studies and U. S. history and government RE each includes three essays # Use of portfolios in all subject areas (grades 4, 8, 12)							
NC	* Portfolio use in reading, math, and language arts (grades 1,2) and for Pre-K and kindergarten use		*				3, 6, 8	Reading, math, language arts
	* Writing tests (grades 6, 8) * Essay writing and geometry proofs are required	*					3, 6, 8	Social studies, science
	◊ Math, science, social studies, writing and language arts (grades 3-8)	*		*	*	HS	8 - 12	Variety
ND			*				3, 6, 8, 11	

* In Place
◊ Planned/Under Development
\# Under discussion

CRT Criterion-Referenced Test
NRT Norm-Referenced Test
Min Comp Minimum Competency Test

STATE	ALTERNATIVE ASSESSMENT PRACTICES	TRADITIONAL ASSESSMENT PRACTICES						
	PERFORMANCE-BASED ACTIVITIES	CRT	NRT	Min Comp	Exit Exam	Level	Grade Level	Test Subject Areas/Comments
OH	* Graduation exams consist of enhanced multiple choice and performance assessments		*				4, 6, 8, 10	Reading, math, language arts
				*	*	HS	9-12	Reading, math, writing, citizenship (class of 1994)
OK	◊ Direct writing assessment (grades 7 and 10 - funding permitting)		*				3, 5, 7, 9, 11	Achievement
		◊			◊	HS	12	Implementation: (1992-1993)
OR	* Enhanced multiple choice assessment in use for reading, math, writing and language arts (grades 3, 5, 8, 11)						8 sample	Reading, math, direct writing
	◊ Direct writing assessment spring 1991 (grades 3,5,8,11)							
	◊ Performance assessments in math, science, social studies, study skills, listening, art, music and health (grades 3, 5, 8, 11)							
	# Exploring writing and math portfolios and interested in physical education, fitness							

* In Place
◊ Planned/Under Development
\# Under Discussion

CRT Criterion-Referenced Test
NRT Norm-Referenced Test
Min Comp Minimum Competency Test

STATE	ALTERNATIVE ASSESSMENT PRACTICES	TRADITIONAL ASSESSMENT PRACTICES						
	PERFORMANCE-BASED ACTIVITIES	CRT	NRT	Min Comp	Exit Exam	Level	Grade Level	Test Subject Areas/Comments
PA	* Direct writing assessment (grades 6, 9) # Science performance testing	*					3, 5, 8	Reading, math
RI	* Direct writing assessment (grades 3, 6) * Merit recognition tests (a voluntary program for grade 12); use both multiple choice and performance testing in over 20 subject areas. ◊ Portfolio use for integrated language arts/math (grade 3)		*				3, 6, 8, 10	Reading, math, language arts
		*					3, 6, 8, 10	Health
			*				3, 6, 8, 10	Physical Fitness
SC	* Direct writing assessment (grades 6, 8, 10) ◊ Performance assessments in reading and math (grades 1-3, 6, 8) and writing (grades 6, 8) # Use of calculators in math testing	*					1, 2	Reading, math
							3	Reading, math, science
							6, 8	Reading, math, science, writing
		*			*	HS		Reading, math, writing
			*				4, 5, 7	Reading, math, science, language, social studies, spelling listening, study skills
							9, 11	Reading, math, language, science, social studies, study skills, spelling
SD			*				4, 8,11	Reading, math, English, science, social studies

* In Place
◊ Planned/Under Development
Under Discussion

CRT Criterion-Referenced Test
NRT Norm-Referenced Test
Min Comp Minimum Competency Test

STATE	ALTERNATIVE ASSESSMENT PRACTICES	TRADITIONAL ASSESSMENT PRACTICES						
	PERFORMANCE-BASED ACTIVITIES	CRT	NRT	Min Comp	Exit Exam	Exit Exam Level	Grade Level	Test Subject Areas/Comments
TN	◊ Writing assessment will be operational for 1991 - 1992 school year (grades 4, 8, 11)	*	*				2 - 8	
			*				10	
				*	*	HS		Math, language arts
TX	* Writing assessment (grades 3, 5, 7, 9, 11)	*		*			3, 5, 7, 9	Reading, math, writing
	◊ Performance testing (maybe integrated writing and social studies)							◊ Social studies, science (1994-95)
		*		*	*	HS	11, 12	Math, English language
								◊ Social studies, science (1994-95)
			*				3 - 11	Reading, writing, math, science, social studies (1991-92)
UT	* Voluntary writing assessments (grades 1-12)		*				5, 8, 11	Achievement
VT	◊ Student portfolio assessment for math and writing to be piloted 1990- 1991, used for reporting purposes 1991 - 1992 and to assess other subject areas in future (grades 4, 8)			*	*		8	Reading, math listening, reasoning (will be optional)
	# Science, history, citizenship							
VA	* Voluntary performance and enhanced multiple choice assessments are embedded in testing in reading, math, writing, science, social studies and language arts (grades K-12) and foreign language (grades 9-12)		*				1	
			*				4, 8, 11	Achievement
					*	8	6	Reading, math, writing

* In Place
◊ Planned/Under Development
\# Under Discussion

CRT Criterion-Referenced Test
NRT Norm-Referenced Test
Min Comp Minimum Competency Test

STATE	ALTERNATIVE ASSESSMENT PRACTICES	TRADITIONAL ASSESSMENT PRACTICES						
	PERFORMANCE-BASED ACTIVITIES	CRT	NRT	Min Comp	Exit Exam	Level	Grade Level	Test Subject Areas/Comments
WA			*				4, 8, 10	Reading, math, language arts Optional: science, social studies
							◊ 11	1991 - to replace grade 10
WV	* Direct writing assessment (grades 8, 10)		*				3, 6, 9, 11	Reading, math, spelling, language arts, reference skills, science, social studies
							3, 9	Cognitive Abilities Test
		◊					1-4 (1991) 5 (1992) 6 (1993) 7 (1994) 8 (1995)	Reading, math, science (spring 1991)
WI	# Science, language arts performance testing		*	*			Usually 3, 7, 10 or once 1 -4, 5 - 8, 9 - 10	<u>Voluntary</u> Reading, math, language arts
							3	Reading
WY	◊ Local districts will use performance assessments in language arts, math and other subject areas in order to meet new accreditation regulations.						8	Pilot math
							4, 8, 12	Periodic funding for NAEP

* In Place
◊ Planned/Under Development
\# Under discussion

CRT Criterion-Referenced Test
NRT Norm-Referenced Test
Min Comp Minimum Competency Test

Performance Testing in New York State's Assessment Programs

New York State's statewide testing programs have already incorporated some performance-based assessment. In fact, one program, a performance-based fourth grade science examination, has been chosen as a model for an upcoming book on alternative assessment. The test, known as a science manipulative skills test, was given to 204,521 students in 1989. The large-scale administration was unique and has garnered significant attention nationally as more states look toward diversifying assessment practices.

In fact, many components of New York State's existing assessment program contain performance testing features. An October 1990 report to the Board of Regents, "A New Compact for Learning: Assessing Results" identifies the following performance assessment activity imbedded into the State's current student assessment program:

ENGLISH: Direct writing assessments are used for the grade 5 Pupil Evaluation Program test, the Preliminary Competency exams given in grades 8 and 9, and all students in grades 11/12 are required to take either the English Regents exam or the Regents Competency exam in writing.

FOREIGN LANGUAGE: Performance writing, speaking and listening are included in the Regents exam for foreign languages and in the second language proficiency exams.

SOCIAL STUDIES: Regents examinations, Regents Competency Tests, and Program Evaluation Tests relating to social studies contain performance testing writing components and optional group participation projects.

MATHEMATICS: All Regents exams and Regents Competency exams in mathematics have segments that "require a pupil to generate responses rather than choosing among alternatives."

SCIENCE: In addition to the previously mentioned manipulative skills test for grade 4 (a Program Evaluation Test), a laboratory skills performance test has been a component of the Regents examination in earth science since 1972. Also, performance and completion of laboratory assignments (completion) are a prerequisite for taking the Regents examination in biology while students taking Regents examinations in biology, chemistry, earth science and physics must generate responses (rather than pick among multiple choice answers).

OCCUPATIONAL EDUCATION: Examinations in business subjects may be totally performance-based (typing and shorthand) or have had performance components (business mathematics, bookkeeping).

ART: Regents diploma students sequencing in visual arts may submit a portfolio of 7 to 10 artwork pieces representing several different media and content areas to demonstrate their competencies.

PHYSICAL EDUCATION: Physical fitness testing has been in place for over 30 years.

NATIONAL PERFORMANCE ASSESSMENT PROJECTS

National Testing Initiatives and Project Development

The activity centered on performance-based testing within individual states is also reflected at the national level. The current national assessment tool, the National Assessment for Educational Progress (NAEP), and another proposed national test targeted for implementation in 10 years will both include the new alternative assessment style of testing. In addition, a growing number of national projects have been announced that will seek to obtain and disseminate information on the development of new alternative assessment strategies and also to assist with implementation problems.

Mandated by Congress in 1969, NAEP has tested samples of students nationwide in a variety of subjects. For the first time, state-by-state comparison data will be made available in mid-1991 based on the 1990 mathematics tests given to samples of eighth grade students in 37 participating states and the District of Columbia. Changes related to alternative testing are also planned for NAEP. The National Assessment Governing Board has come under pressure by a coalition, led by FairTest, to make changes in the assessment format of the NAEP tests. In late March 1990, the Board announced the adoption of changes to be implemented in the 1992 NAEP reading and writing assessments. Responding to calls for inclusion of alternative assessment formats, increased use of open-ended questions will now be used to test higher-order skills. Open-ended test items will increase from 20 to 40 percent of the test questions. In addition, pilot studies are under way assessing oral reading abilities and writing portfolio use. NAEP had been strongly pressured by FairTest and some 75 organizations to focus more resources on development of alternatives to multiple choice tests. Even with the latest NAEP efforts, this group of testing reform advocates concludes that the movement is too slow. Significant organizations supporting this position include: the American Association of School Administrators, the National Parent-Teachers Association, the National Association for the Advancement of Colored People, the National Education Association, the National Association of Secondary School Principals, the Association for Supervision and Curriculum Development, the National Council of Teachers of English, the American Reading Council, the National Coalition of Advocates of Children, the National Council of the Social Studies, and the International Reading Association.

Development of a new national examination system that would turn to other than multiple choice items has received initial funding of $1.3 million from the John D. and Catherine T. MacArthur Foundation. However, the release of this funding hinged on the project attracting an additional $1.1 million. In January 1990, the Pew Charitable Trusts of Philadelphia provided this amount. The awards will fund a joint project to be conducted by the National Center on Education and the Economy and the University of Pittsburgh's Learning Research and Development Center. The new assessment system

would include three components: performance examinations, portfolios and projects. Lynn Olson of **Education Week** reports that such national level exams "eventually could be used as the basis for high school diplomas, college admissions, and employment." (December 12, 1990, p. 5) Students aged 14-16 would be tested with provisions included for multiple retesting opportunities for those failing. Piloting of two subjects would take place between January 1991 and June 1992. Widespread use for all core subjects (reading, mathematics, writing, science, history/geography/civics and work readiness) would occur by 1997 and full implementation by 2000.

Interest in alternative assessment has become significant enough to result in the formation of new projects by national groups to explore the development of new testing mechanisms as alternatives to standardized tests. Among the entities engaged in such projects are the Council of Chief State School Officers (CCSSO), the Urban District Leadership Consortium (UDLC), the Education Commission of the States (ECS), UCLA's Center for Research, Evaluation, Standards and Student Testing (CRESST), and the U.S. Congress Office of Technology Assessment (OTA). In some cases, the projects are collaborative efforts with many diverse partnerships. The purpose of these projects is to identify current and planned research for future reference as new assessment plans are devised for large-scale application.

Urban District Leadership Consortium Projects and Partnerships

The Urban District Leadership Consortium Assessment Projects are affiliated with the American Federation of Teachers Center for Restructuring in Washington, D.C. The Consortium is an organization of 25 urban districts interested in reform and collaboration on several projects that focus on alternative assessments and accountability. UDLC has three separate projects under way or in the planning stage. Each project is focused at a different level of the education process--the classroom, the school and the district.

The first project is identified as "in the classroom," and expects to train 390 high school science and mathematics teachers in performance assessments over a two-year grant period (October 1, 1990 to September 30, 1992). The project will use performance tasks developed by the Connecticut State Department of Education, including 32 laboratory-based science tasks and 18 mathematics tasks. The mathematics tasks have been developed as "assessment prototypes" that are "all sustained problems with either multiple correct solutions or several alternative solution paths." The project proposal sets the following five goals:

(1) providing training to urban and rural mathematics and science teachers in different aspects of performance assessment strategies;
(2) exploring the teaching and learning implications of extended performance assessments in school restructuring efforts;

(3) determining the feasibility of using performance assessment strategies in inner city and rural schools by identifying the implementation problems and the potential solutions;
(4) developing feasible procedures for scoring complex assessment tasks, and establishing performance standards and reporting formats for use by high school science and mathematics teachers and policymakers; and
(5) developing a series of training papers and a videotape library of models.

This particular project is a joint effort that has the UDLC collaborating with both the Connecticut State Department of Education and ECS's Project Re: Learning Consortium. Funding for the project was provided by the National Science Foundation. Other funding is being sought to expand the project by including a training component and establishing a telecommunications network for those UDLC teachers involved in the project. The sites are in 18 UDLC districts nationwide, including Rochester, New York. (See Appendix G for complete list of UDLC participants.)

The second UDLC assessment project is school-centered and is a collaborative effort with Boston College's Center for the Study of Testing, Evaluation and Educational Policy which is now seeking funding for pilot projects. The eight districts that have made a commitment to participate in the three- to four-year study are: Boston; Chicago; Dade County, Florida; Kansas City, Missouri; Los Angeles; Philadelphia; Rochester; and San Francisco. The project will examine the process used by the district and the schools to develop districtwide performance-based assessments. The choice of school-generated indicators to meet both the district goals and the special needs and circumstances reflected in each school's own goals will be studied. John Cawthorne, Project Director, expects that the performance-based assessments will be for reading, writing and mathematics and, "ideally," will be implemented in the pilot schools at grades 2,5,8,10, and 12. Once funded, the project schedule calls for preliminary meetings in the spring and summer with first year pilot school programs under way in the 1991-92 school year.

The third UDLC project is district-centered. Four "working conversations" are planned to examine alternative assessment use in large urban districts. The first meeting was held in Philadelphia in August, 1990 where the discussion centered on the following themes:

- alternative assessment development must be accompanied by restructuring of curriculum instruction and governance;
- the barriers to performance assessment in an urban setting include costs, regulations--Chapter I and court-ordered desegregation guidelines, and the lack of teacher training and support;
- alternative assessment has some legitimate and unresolved questions of reliability and validity; and

- new broader concepts of accountability need to be developed for the use of alternative assessments to avoid the undesirable narrow use for accountability purposes now seen in many places with standardized testing.

The forthcoming meetings will expand on these ideas with the other participant districts (also listed in Appendix G). Consultants for this project are affiliated with Teachers College/Columbia University, Harvard, and Boston College.

Council of Chief State School Officers' Project

In December 1990, the CCSSO's State Education Assessment Center released details of the newly established Interstate Consortium on Alternative Methods of Student Assessment that is geared to assist states in adopting improved methods of student assessment. The collaborative effort will involve state departments of education, the U.S. Department of Education's Office of Educational Research and Improvement (OERI), CRESST at UCLA, and the Education Commission of the States (ECS). The project is receiving funding from the U.S. Department of Labor and the John D. and Catherine T. MacArthur Foundation.

The Interstate Consortium project will focus on two separate activities. The first, a laboratory to develop alternative methods of assessment, will assist states with technical issues identified through the following questions:

- To what extent can alternative forms of assessment improve the validity of large-scale student assessments in relation to desired student knowledge and skills?
- Can sufficient reliability of scores from alternative methods of assessment be maintained as new exercise formats are used and student results are couched in fewer, more complex tasks?
- Is there positive benefit in the trade-off of gains in validity against losses in reliability?
- How can we score new kinds of exercises effectively and consistently?
- How can assessment (including new forms of assessment) be presented to teachers and school systems so as to avoid undesired focusing of instruction?

Additional areas of interest that are expected to be addressed relate to the cost factor. The Interstate Consortium laboratory component anticipates assisting in the development of "economically efficient" techniques for scoring the new performance-based assessment activities. The Consortium will seek to provide help with the overall analysis of costs and feasibility of alternative assessment since, to date, estimation of costs for the newer

assessment practices has proven to be "extremely complex." CCSSO will gather data on each type of assessment, and after analysis, provide states with the results.

Another side of the Consortium's laboratory function calls for the development of an assessment information exchange for states. In addition, the Consortium will serve as broker for states to receive technical assistance with developing performance style exercises, conducting reliability studies related to administration and scoring, and applying research on cognition to the design of new exercises.

The second major function of the Consortium is to establish an item pool for alternative methods of student assessment as a means to help states to build on the work and success of other states and, by sharing test items, cut costs and enlarge the number of the testing items available. Large-scale assessments require numerous items to insure reliable results. Development of the new type performance items requires time, creativity and cost, thereby, making item pooling an attractive and logical action. However, legal and proprietary rights issues must be addressed prior to such action. Development of the item pool is expected to initially involve 15 to 20 states who will develop a classification scheme for alternative assessments and types of items and define standards for evaluating items.

CRESST at University of California at Los Angeles

Federal research monies will support multiple projects in alternative/performance testing at UCLA's Center for Research, Evaluation, Standards, and Testing over the next five years. In November 1990, $14.3 million were awarded for this purpose with the primary focus of the research on performance-based testing. CRESST will develop projects related to disseminating information on the new assessment activities, working out the psychometric issues, and assisting with the problems associated with implementation.

As one partner in the CCSSO Consortium, CRESST will devote attention to the clearinghouse function. Information on performance testing activities will be collected and made available to state legislative policymakers and staff, State Education Agency staff and interested organizations. A technical issues hotline is planned as well as a directory of resource personnel to help with performance assessment problems. The psychometric issues of reliability and validity related to performance-based testing will also be undertaken. In addition, CRESST will also become directly involved with assisting in the improvement of the performance-based programs that states get up and running (such as in Connecticut and Vermont). An abstract detailing these projects is expected to be available by early 1991.

Office of Technology Assessment - U.S. Congress

The House Committee on Education and Labor and the Senate Committee on Labor and Human Resources have requested the Congress Office of Technology Assessment (OTA) to undertake a comprehensive assessment of alternative measurement technologies. The main focus of the project is "to examine the technological and institutional features of new measures being developed and evaluated. A focal point of the study will be the appropriateness of various technologies for diagnosing individual learning needs, measuring skills acquired in school and their relation to future education or employment, and developing performance indicators for accountability of schools and school systems. Comparisons with other countries' testing strategies are anticipated as well." (p. i) In response, OTA is in the process of developing three separate projects to gather the information needed for the report to the U.S. Congress. That report, scheduled to be ready by September 1991, has the working title, **Schools, Kids, and Measurement: Technologies of Assessments.** The project proposal indicates that, in addition to exploring performance-based testing and an international comparison on assessment practice, OTA will include components in the study on computer-based technologies that analyze and diagnose learning patterns and cognitive styles and examine use of nontest-based indicators (attendance, dropout rates, etc.) Details on this last project are not yet available.

As of November 1990, the performance-based assessment component was under way as Ruth Mitchell, an alternative assessment consultant to OTA, has begun working on case studies involving six alternative assessment programs. One case study will explore the state-level alternative assessment program in Arizona. The other case studies focus on examining:

- school-based elementary art assessment (Bozeman, Montana);
- district-level oral presentation competition as part of the "We The People" national competition (Birmingham, Alabama);
- multiple measures of assessment including teacher portfolio use (South Brunswick, New Jersey);
- district testing that exceeds the California Assessment Program (San Diego); and
- advanced placement examination use for studio art assessment.

Details on other aspects of this project have not been decided, although the project schedule indicates that two to three workshops will be convened during the study to bring together test publishers, cognitive scientists and researchers, education personnel from the states, and teachers. The case studies and reports from contractors are intended to provide background information for the numerous assessment questions outlined in the study proposal. Among the questions designated to be addressed by the OTA are the following related to alternative assessment:

- What are the probable effects of different kinds of performance assessments on teaching style and curriculum design?
- What special training do teachers need in order to adopt alternative assessment techniques?
- Can alternative assessments be applied to various objectives of testing, such as diagnosis of individual learning needs as well as school or district-level performance?
- Can alternative assessments be used for evaluation of Federal compensatory education programs such as Chapter I?
- What are the funding requirements for continuing research on alternative assessment methods?
- Can innovative performance-based assessments remove cultural, racial, ethnic, and gender bias alleged to exist in multiple choice tests? Or, will their necessarily more subjective scoring only compound these problems?

One major concern related to the new testing is the cost factor. OTA raises concerns about the testing industry's role and questions "whether large firms that dominate the testing market will exploit their financial advantages by investing in innovations or whether they will prefer to maintain the status quo." With this in mind, OTA will respond to the following issues in their 1991 project:

- How much is currently spent on testing in American schools, and how might that change with new test instruments? Do estimates of spending account for indirect and opportunity costs?
- How does the structure of the testing market influence innovation and implementation of alternative assessments?
- How can the benefits of alternative testing programs be measured in terms of their usefulness to students and school systems?

Test publishers have recognized the potential of the new assessment techniques. They have prominent roles in the development of the new state-level performance testing in Maryland and Arizona. The Educational Testing Service (ETS) has announced plans to introduce new performanced-based instructional examinations for classroom use. Noting that testing is in the process of changing, ETS contends that standardized testing should not be totally discarded but could be complemented by the new forms of assessments.

TESTING AROUND THE WORLD

Research Availability

As the interest in moving beyond standardized testing and existing practices has increased, there has been a growing interest in how other countries test their students. Other researchers interested in the new trends towards performance-based testing, including Ruth Mitchell of the Council on Basic Education, Nancy Mead of ETS and Fred Frelow of Columbia's Teachers College, have indicated that while such information would be valuable, they found that there is scant international research material currently available that directly relates to this assessment issue. While in telephone interviews assessment researchers/consultants generalize about the tradition of essay examinations in European countries, no comprehensive comparison of the assessment practices of other countries is available. References to performance-based activities in other countries have been included in some news articles relating to the new alternative assessments, but specifics are generally lacking.

The Congressional OTA has plans to undertake such a comparative survey as a component of the previously detailed project on **Schools, Kids, and Measurement: Technologies of Assessment**. As of November 1990, OTA has not decided whether that office or a hired consultant will collect the data on foreign countries. The project proposal identifies the following issues to be examined in this particular component of the study:

- What can be learned from the experiences of other countries about the relationship among testing, curriculum design, and teaching?
- What are the most appropriate tests to use for international comparisons of educational performance?

Canadian Provincial Assessment Activity

The most up-to-date information on performance assessment activity outside the United States relates to three provinces of Canada. The 1990 Michigan Department of Education's survey of large-scale assessment programs detailed performance assessment plans in British Columbia. Information on Ontario and Alberta's Provincial assessment programs was received from provincial and test publisher contacts in response to OPRSS solicitation of materials. A brief summary of the most relevant information follows.

Performance-based testing being developed in British Columbia currently has two major testing programs, the Grade 12 Examinations Program and the Provincial Learning

Assessment Program. The Grade 12 examinations are given to all students taking the following courses: English 12, Algebra 12, History 12, Literature 12, German 12, French 12, Latin 12, Physics 12, Chemistry 12, Biology 12, and Communications 12. The Provincial Learning Assessment Program assesses students in grades 4, 7, and 10 in either mathematics, science, language arts, or social studies. One subject is chosen each year for testing. Both examination programs include multiple choice and "extended response" test items with plans to expand the tests to include performance testing components. The Learning Assessment test program expects to use trained observers/assessors of performance tasks on a sample of students. Implementation schedule and further details were not provided.

Information received from Nelson Canada's Measurement and Guidance Division provides insight into assessment activity in the Province of Ontario. Reporting on the topics of a December meeting of the Association of Educational Research Officers of Ontario, Nancy Power of Nelson Canada confirms that the trend in Ontario is away from the standardized testing and toward performance-based testing. Power, in a letter to OPRSS, provides the following commentary on the current assessment activity in Ontario:

> The Ontario Ministry presently conducts Provincial Previews not only because they are accountable to the public but to provide valuable insights into the present programs. The Ministry does however recognize that an assessment is more valid if it measures objectives the same way as they are taught. The 1989 and 1990 Math tests for students in grades 6 and 8 required the use of manipulative and the demonstration to the test administrator of how they arrived at their final answer (in addition to multiple choice questions). Similarly, the writing test for grade 12 presently being developed will reflect the present method of instruction which focuses on the 'writing process.' Not only will assigned writing tasks be evaluated but students will also be required to select their best piece of writing they completed during that school year. This sample will be evaluated by the administrators of the test and in addition, students will be asked to explain why they think it is their best sample. From this self-evaluation the administrators can learn what criteria the students themselves use to evaluate writing. Finally, students will submit their writing portfolios which contain a collection of their written work from throughout the year. Given the amount of work required for this assessment, only a sample of students will be required to participate. This is for the Ministry's use and is intended to provide feedback on the system as a whole.

Power also describes two classroom-oriented performance assessment programs. The Ministry is considering a Benchmark program which will provide schools with learning outcome definitions that are meant to set the criteria but not be a method of assessment in itself. Toronto is implementing an assessment program with the same name which provides teachers with criteria of performance. Nelson Canada is working with a school district, Etobicoke, in developing a cross-curricular, performance-based assessment program patterned after the English "Touchstones" program. Adapted from the English version, the Ontario program will use a "test as we teach" approach that is "activity based, involve[s] both group and individual work, and provide[s] open-ended questions to encourage problem solving, investigation, estimation, and free thinking."

In Alberta, provincial examinations are given to private and public school students. Students in Grade 12 take "diploma" examinations which are based on Alberta curriculum specifications in English, social studies, mathematics, biology, chemistry, physics and French. In order to graduate from high school, students must pass the two-part diploma examinations in English which consist of multiple choice and written components. The examinations in biology and physics are 80 percent multiple choice and 20 percent written responses, while the mathematics examination is 80 percent multiple choice and 20 percent written responses. Calculator use is permitted for the mathematics, chemistry and physics examinations. In addition, Alberta tests students in grades 3, 6, and 9 each year. Subjects tested are language arts (English and French), social studies, mathematics and science. Only one subject is tested at each grade level per year on a 4-year rotation schedule. For example, science might be tested for third graders in 1991 with previous science testing administered for this level in 1987 and 1983. The tests for this achievement testing program are criterion-referenced (tied to the Alberta curriculum) and consist of both multiple choice and written response items. At each grade level tested, a direct writing assessment is conducted in language arts. These tests provide both individual and group results and may not be used for promotion or placement decisions.

Literature Review Findings

A limited amount of pertinent materials on assessment in other countries is readily available. Among the literature reviewed, only five articles provided insight into the testing activity taking place outside the United States. Burstall (1986) describes performance-based testing in the United Kingdom; Postlethwaite (1986) details standardized test use in four European countries; Eckstein and Noah (1989) examine end-of-school exiting examinations in eight countries (including the United States). Articles by Voss (1985) describing the oral component of the West German exiting examination and van den Bergh (1987) detailing a feasibility study of "oracy" (listening and speaking) performance testing conducted in the Netherlands were also reviewed, but were of lesser value.

The United Kingdom's national assessment program, the Annual Performance Unit (APU), uses performance testing strategies to assess public and nonpublic school students in England, Wales and Northern Ireland. The APU program tests samples of primary and secondary students (by age rather than grade level) as follows: mathematics and English language -- ages 11 and 15; French, German and Spanish -- age 13, and science -- ages 11, 13, and 15. The APU test results are meant to provide information on the national condition of education. In addition to written tests, the APU also uses "practical" testing that expects the student to demonstrate mastery. The practical testing is performance-based thereby assessing "what students can actually do, not just what they can express in writing." Specially trained teacher-assessors give the untimed tests. The mathematics practical test is in an interview format and the scoring includes an attitude assessment component. The student responds to oral questions and a discussion of mathematics concepts between the assessor and student may occur. Both the English and foreign language practical tests center on "oracy" or "the ability to communicate effectively in the spoken, as opposed to the written, medium." Such tests examine the listening and speaking skills through realistic testing tasks that include instructing and directing; giving and interpreting information; narrating; describing or specifying; and discussing for collaborative purposes or to justify an individual point of view. Audio-cassettes are used for later scoring, although an "on-the-spot impression rating" is also given. Finally, the science practical test aims "to assess the student's ability to carry out investigations, to observe and record skillfully, to generate hypotheses, and to use measuring instruments and other scientific apparatus." The important role played by the teacher-assessor, and of the collaborative expertise of experienced teachers in development of these practical assessments, is highlighted by Burstall. A high level of teacher enthusiasm is reported for participating teachers.

Standardized test use in England, Sweden, the Netherlands, and West Germany is discussed by Postlethwaite. Standardized tests are defined by this author as "any tests that are generated outside the school and that are administered in a common fashion to students in a variety of schools." England utilizes such standardized test for special needs assessment and for guidance purposes. English standardized tests are generally achievement ("attainment") tests or ability ("intelligence") tests, with such use greater in the primary grades than in secondary schools. As of 1986, the Federal Republic of Germany had no nationwide standardized tests. Intelligence and diagnostic standardized tests are used by trained personnel only for guidance/career purposes. Achievement tests are available from textbook publishers but there are little data on their use. Postlethwaite concludes that standardized tests are rarely used in German schools and there is no established culture of such testing in schools. The Netherlands has national examinations that combine internal assessment and the national written examination. Those criterion-referenced tests that are available in certain subject areas are intended for use by teachers for "formative education" and diagnostic use but such use is not significant. Finally, of the countries studied, Sweden makes the most use of standardized testing (for diagnostic and achievement purposes). In some cases the tests are voluntary, but there is a required use

as ordered by the [Swedish] National Board of Education for national assessment purposes.

Eckstein and Noah provide insight into the end-of-secondary schools examinations in England, France, Japan, China, West Germany, Sweden, the Soviet Union and the United States. The examinations described in many cases were the national examinations that secondary school students take to determine the next step in their future schooling, trade or employment. Multiple choice, machine-scored tests were used only by the United States and Japan. (The U.S. tests examined are described as privately-developed tests that are taken "by large numbers of candidates for higher education institutions"--obviously the SAT/ACT examinations.) The exams in all other countries (England, France, China, West Germany, Sweden and the Soviet Union) employed long or short essays, step-by-step computations or explanations, diagrams and similar types of responses. The Soviet Union continues a traditional use of oral examinations, while West Germany is moving away from the practice. In most other countries, oral examinations were rarely used except in the case of foreign language assessment.

International Assessment of Educational Progress

The International Assessment of Educational Progress (IAEP) is a study of 9 and 13 year-old student samples which is planned for administration in March 1991 in 20 nations. In the United States the IAEP will be under the direction of ETS with funding provided by the National Science Foundation and the U.S. Department of Education for the multiple choice component of the testing process. A performance-based testing component is also planned, but funding for United States participation was denied by the funding agencies who reportedly feel that performance assessment is "still in the developmental stage" and therefore inappropriate for "such a high-profile comparison."

The Educational Testing Service field-tested the performance test components which were developed in England and Scotland. Students will be required to conduct experiments and measurement tasks. Nations that will participate in the performance component are: England, Scotland, Ireland, Hungary, the Soviet Union, China, Taiwan, Korea, Mozambique, Canada and Switzerland. In addition to the United States, the countries that will not participate in the performance component are: Italy, France, Spain, Portugal, Jordan, Israel, Brazil and Slovenia (Yugoslavia).

CONCLUSION

The accountability and school restructuring elements of the present search for education reform have contributed to renewed interest in student assessment practices. Testing is an important component of today's schooling, yet states vary in the amount and type of testing in use. Numerous states are in the process of making changes to their testing programs. Some are looking at alternative assessment as a move away from a strictly multiple choice format in an effort to strengthen their curricula now limited by the score-sheet bubble of the standardized test. New York's proposed **A New Compact for Learning** incorporates performance-based assessments by proposing the use of student portfolios, teacher evaluations, and acknowledging that other examinations need not be of the multiple choice variety. In addition, there are elements of performance-based assessment included in many of the state tests currently in use.

Test reform activists have initiated a campaign to limit the mass testing of young children in New York. While state testing programs do not test younger students (under age eight), FairTest and NYPIRG have requested the Board of Regents and the Commissioner of Education to ban such local testing by the Fall 1996. Nationally, concern has been voiced that overtesting is occurring in some states where state testing programs are accompanied by significant local testing practices and mandated testing is required for participation in Federal education programs such as Chapter I. The Compact calls for a review of this existing extensive, seven-part State testing program. To undertake such action would be responsive to the concerns expressed by test reform advocates who are opposed to overtesting of students.

Undertaking a review of the existing New York State program also demonstrates a leadership role in moving ahead with testing reform. Such a review will reveal that some the tests currently in place already incorporate some alternative assessment type formats. In fact, the fourth grade science program evaluation test's manipulative skills component has received national recognition as a model program. New York stands with California, Connecticut, Arizona, Maryland and Vermont as states seeking to move forward with innovative testing activities. In general, New York State's assessment programs and proposals are found to be in harmony with the desirable testing practices identified.

The development of performance-based assessment for large-scale use is now under way throughout the nation. Skeptics worry that alternative assessments are being asked to accomplish too much and that the new test will run the same risk of the very problems associated with "high stakes" standardized testing. Acknowledging that there are some problems associated with multiple choice use and abuse, supporters urge that these problems be addressed and that multiple choice testing not be abandoned. Advocates for new assessment practices are convinced that the improved curriculum and teacher involvement will be important steps in the long-awaited restructuring of education. While

there are concerns raised that the new assessment is moving too fast, the proponents argue that in order to determine the real worth, performance-based testing must be initiated on a large-scale.

Several national projects have been organized to seek answers to some of the unknowns related to implementation of the new tests. With the very bleak fiscal climate, the most significant problem seems to be the increased costs (now estimated by ETS at $50 per portfolio evaluation). Foreign countries have historically used performance testing methods and, while comparative literature is presently scant, their experiences can provide information on the implementation strategies and technical concerns. OTA will examine other countries as part of a planned study scheduled for completion in September 1991.

In conclusion, New York State's existing and proposed student assessment activities are clearly part of the current trend toward more innovative assessment practices. The **Compact for Learning** assessment proposals to include portfolios and to review and re-evaluate the testing now in place demonstrates a willingness to explore further change--action that is sensitive to turmoil now surrounding student testing practices and their use.

APPENDIX A

NEW YORK STATE ASSESSMENT PROGRAMS: DESCRIPTIONS AND SCHEDULING

Pupil Evaluation Program (PEP): Reading and mathematics tests, grades 3 and 6; writing test, grade 5. Annual spring administration; all pupils, public and nonpublic. Purposes: (1) identification of pupils needing remediation, which is mandatory; and (2) school assessment.

Program Evaluation Tests: Social studies, grades 6 and 8; science, grade 4. Annual spring administration; all pupils, public and nonpublic. Purpose: school assessment.

Preliminary Competency Tests: Reading and writing, grades 8 or 9. Annual administration to pupils scoring below median on last PEP test in respective subject. Primary purpose: identification of pupils needing remediation, which is mandatory.

Regents Competency Tests: Reading and writing, grades 11 or 12; mathematics, grades 9-12; science, grade 9; global studies, grade 10; United States history and government, grade 11. Offered under secure conditions in January, June and August. Purpose: demonstration of competencies required for local high school diploma.

Second Language Proficiency Examinations: French, German, Italian, Latin and Spanish, grades 7, 8 or 9. Purpose: to permit students to earn high school credit for the study of a second language in elementary and middle schools.

Occupational Education Proficiency Examination: Achievement tests in Introduction to Occupations, technology, home economics, health occupations, agriculture and business, grades 9-12. Offered under secure conditions in January, June and August. Purposes: (1) establish and maintain standards in occupational education; (2) demonstration by occupational education majors of competencies required for high school diploma; and (3) marking and grading.

Regents Examinations: Achievement tests in 17 high school subjects. Offered under secure conditions in January, June and August. Purposes: (1) marking and grading; (2) eligibility for Regents diploma; and (3) demonstration of competencies required for local diploma.

Test Development: Reading tests (Degrees of Reading Power) purchased from the Touchstone Appied Science Associates; all other tests developed in-house with cooperation of classroom teachers.

Source: New York State Education Department, Division of Educational Testing; March, 1991.

APPENDIX A (continued)

New York State Testing Program Components and Grades at Which Tests Are Administered

Components	Grade Administered									
	3	4	5	6	7	8	9	10	11	12
Pupil Evaluation Program*										
Reading	x			x						
Mathematics	x			x						
Writing			x							
Program Evaluation Tests*										
Science		x								
Social Studies				x		x				
Preliminary Competency Tests**										
Reading						x	x			
Writing						x	x			
Regents Competency Tests**										
Reading									x	
Writing									x	
Mathematics							x			
Science							x			
Social Studies								x	x	
Regents Examinations**							x	x	x	x
Second Language Proficiency Examinations**						x				
Occupational Education Proficiency Examinations**							x	x	x	x

*All students tested at grade indicated

**Selected students tested

Source: New York State Education Department. Report to the Board of Regents: "A New Compact for Learning: Assessing Results." October, 1990.

APPENDIX B

NORM-REFERENCED AND CRITERION-REFERENCED MEASUREMENT

Basic Terminology

Norm-Referenced Test	A test designed to provide a measure of performance that is interpretable in terms of an individual's relative standing in some known group.
Criterion-Referenced Test	A test designed to provide a measure of performance that is interpretable in terms of a clearly defined and delimited domain of learning tasks.*

Comparison of Norm-Referenced Tests (NRTs) and Criterion-Referenced Tests (CRTs)**

Common Characteristics of NRTs and CRTs:

1. Require specification of the achievement domain to be measured.
2. Require a relevant and representative sample of test items.
3. Use the same types of test items.
4. Use the same rules for item writing (except for item difficulty).

5. Are judged by the same qualities of goodness (validity and reliability).

6. Are useful in educational measurement.

Differences Between NRTs and CRTs (only a matter of emphasis):

1. NRT - Typically covers a large domain of learning tasks, with just a few items measuring each specific task.
   CRT - Typically focuses on a delimited domain of learning tasks, with a relatively large number of items measuring each specific task.

2. NRT - Emphasizes discrimination among individuals in terms of relative level of learning.
   CRT - Emphasizes description of what learning tasks individuals can and cannot perform.

3. NRT - Favors items of average difficulty and typically omits easy items.
   CRT - Matches item difficulty to learning tasks, without altering item difficulty or omitting easy items.

4. NRT - Used primarily, but not exclusively, for survey testing.
   CRT - Used primarily, but not exclusively, for mastery testing.

5. NRT - Interpretation requires a clearly defined group.
   CRT - Interpretation requires a clearly defined and delimited achievement domain.

Source: Norman E. Gronlund. Measurement and Evaluation in Teaching (95th Edition, 1985).

---

* New York State's testing program uses criterion-referenced tests that measure a sample of the content domain in the State syllabi.

**When each is built to maximize its type of interpretation.

APPENDIX C

TYPES OF QUESTIONS CONTAINED IN THE NEW YORK STATE TESTING PROGRAM

	Percent of Questions		
		Open Ended*	
	Multiple Choice	Paper and Pencil	Other
Pupil Evaluation Program	80%	20%	-
Program Evaluation Tests	74	15	11
Preliminary Competency Tests	50	50	-
Regents Competency Tests	68	32	-
Regents Examinations	61	25	14
Second Language Proficiency	62	11	27
Occupational Education	100	-	

*Includes essay questions, performance tests, and completion items.

In general, 80 to 90 percent of the multiple-choice questions and virtually all open-ended or essay questions used in the State testing program measure higher order thinking skills, such as the ability to synthesize, evaluate, and analyze information.

Source: New York State Education Department. Report to the Board of Regents: "A New Compact for Learning: Assessing Results." October, 1990.

APPENDIX D

ADVOCATES FOR REPLACING STANDARDIZED TESTS WITH ALTERNATIVE TESTING PRACTICES

K-12 Testing
Signers of "Statement on Genuine Accountability"

The following organizations and individuals have signed the "Statement on Genuine Accountability:"

American Association of Colleges for Teacher Education

American Federation of Teachers

American Reading Council

Angelo N. Ancheta, Asian Law Alliance*

APPLE Corps, Atlanta, GA

Association for Supervision and Curriculum Development

Association for Women in Science

Bank Street College of Education

Rims Barber, Mississippi Human Services Agenda

Leonard Beckum, City College of City University of New York

Diana Caballero, Puerto Rican/Latino Education Roundtable, NYC*

Center for Law and Social Policy

Center for Women Policy Studies

Council for Basic Education

Harold E. Dent, Psychological and Human Resources Counsultants*

Education Law Center, Newark, NJ

Pablo Eisenberg, Center for Community Change*

Federation of Organizations for Professional Women

Fund for the Feminist Majority, Eleanor Smeal, President

Howard Gardner, Project Zero, Harvard University*

Genesee Valley Developmental Learning Group, New York

La Donna Harris, Americans for Indian Opportunity*

Leslie A. Hart, Brain-Compatible Education Associates

Asa Hilliard, Georgia State University*

*Holt Associates/Growing Without Schooling

Institute for Learning and Teaching, St. Paul, MN

International Reading Association

KEY-Kids, Education and You, Denver City, TX

Massachusetts Advocacy Center

Matsushita Foundation, Inc.

Deborah Meier, Central Park East Secondary School, NYC*

Samuel J. Meisels, University of Michigan*

Sara E. Melèndez, Vice-Provost, University of Bridgeport*

Mexican American Women's National Association

National Association for the Advancement of Colored People

National Association for the Education of Young Children

National Association of Elementary School Principals

National Center for Fair & Open Testing (FairTest)

National Coalition of Advocates for Students

National Coalition of Title I/Chapter I Parents

National Council of Teachers of English

National Indian Youth Council

National Organization for Women-NYC Chapter

National Parent-Teacher Association

National School Boards Association

National Women's Law Center

National Women's Political Caucus

Fred M. Newmann, National Center on Effective Secondary Schools*

New York Public Interest Research Group (NYPIRG)

Vito Perrone, Harvard University*

Project Equality Inc., Kansas City, MO

Puerto Rican Legal Defense and Education Fund

Rochester (NY) Teachers Association

Lori Rubenstein, Partnership for Democracy, Washington, DC*

William V. Schipper

Donald H. Smith, Baruch College*

Southern Association on Children Under Six

Southern Christian Leadership Conference

Southern Regional Council

Gail E. Thomas, Texas A & M University*

United Federation of Teachers (NYC)

United States Student Association

Dorothy J. Watson, Whole Language Umbrella*

Paul Weckstein, Center for Law and Education*

Grant Wiggins, CLASS

Arthur E. Wise

Women's Research and Education Institutes

*Organization listed for identification purposes only

Source: The Center for Fair and Open Testing, **Fair Test Examiner**, Winter 1989-1990, p. 7; Summer 1990, p. 6.

APPENDIX E

ADVOCATES FOR BAN ON STANDARDIZED TESTING OF YOUNG CHILDREN

100 Black Men; Larry Dais, Chairman
Advocates for African American Children; Georgia McMurray, President
Advocates for Children; Norma Rollins, Director
Alliance for Parental Involvement in Education, Inc., Katharine Houk/Seth Rockmuller
Arise Research Bureau, Mya Savaant, Director of Research and Development
Association for Childhood Education International; Mary McKnight-Taylor, President
BCC Child Development Center, New York City; Elba Velez, FDC Coordinator
Beauty and the Beast Storytellers; Martha Hamilton, Mitch Weiss
Black Task Force on Child Abuse and Neglect; Rosalyn Ferguson, Chairperson
Bronx New School, Parent Teacher Association; Janitzia O'Neil, Parent Coordinator
Brooklyn New School, New York City; Paul Schwarz, Director
Buffalo State College Department of Elementary Education and Reading; Graduate Reading Faculty
Capital District Association for the Education of Young Children; Mary Anne Fratelli, President
Center for Collaborative Education; Heather Lewis, Director
Champlain Valley Association for the Education of Young Children; Cynthia Haley, President
Child Care, Inc.; Antony Ward, Executive Director
Diapering Alternatives; Melanie Lyte Peterson, Coordinator
Early Childhood Education Council of New York City; Nina Lublin, President
Early Childhood Education Council of Western New York; Jeanne Petersen, President
Genesee Valley Developmental Learning Group; Ardis Tucker, Chairperson
NAACP Legal Defense and Educational Fund, Inc.; Julius Chambers and Norman Chachkin
New Program, P.S. 261; Anne Powers
New York INTERFACE Development Project, Inc.; Marjorie Wilkes, Co-Director
New York State Association for the Education of Young Children; Fred Barnes, President
New York State Child Care Coordinating Council; Sandra Lamm, Director
New York Urban Coalition; Glenn Pasanen, Director of Public Policy
North Country Association for the Education of Young Children; Linda Pollock, President
Parent Association of P.S. 19Q; Dora Cuadra, President
Parents Coalition for Education in New York City, Inc., Claudia Butler, Executive Director
People About Changing Education (PACE); Don Murphy
Public Education Association; Shelia Haber
Puerto Rican Legal Defense and Education Fund
Primrose School of Lincolndale, New York; James Beaty, Principal
Ravena Coeymans Selkirk Schools; Nancy Andress, Director of Instructional Services
Richard C. Owen Publishers, Inc.; Richard C. Owen, President
River East School, New York City; Laura Daigen, Director
Rockland Council for Young Children; Judith Savage, Assistant Director
Syracuse Association for the Education of Young Children; Nancy Jenner Gabriel, President
Westchester Mid-Hudson PreKindergarten Directors; Jeanne Millman
United Parents Association of New York City; Jan Atwell, President

Source: New York Public Interest Research Group, September, 1990.

APPENDIX F

DATA SOURCES FOR 50 STATE TESTING COMPARISON CHART

This study's 50 state testing comparison chart identifies the practices that are in place, planned/under development, or under discussion based on information provided from the following sources:

- **Descriptions of Performance Assessment in the States** (Council of Chief State School Officers, December 1990);
- **Performance Assessments in the States** (Council of Chief State School Officers, December 17, 1990);
- **State Policies and State Role in Curriculum** (Council of Chief State School Officers, May 1990);
- **Educational Standards in the 50 States: 1990** (Educational Testing Service Policy Information Center, August 1990);
- **Association of State Assessment Programs Survey of Large-Scale Assessment Programs** (Michigan Educational Assessment Program, April 1990); and
- **CRESST Survey: State Activity and Interest in Performance-Based Assessment** (Center for Research, Evaluation, Standards and Student Testing at UCLA, June 1990).

In addition, the Office of Planning, Research and Support Services (OPRSS) solicited information on the status of assessment programs in the states, certain Canadian provinces and from a variety of other entities. When appropriate, such information was incorporated into the Chart on Alternative and Traditional Assessment Practices in the United States--February 1991. Also, the ECS Clearinghouse provided information on 1990 State legislative action related to assessment. Other materials prepared by the State Education Department's assessment staff for presentation to the Board of Regents in October 1990 (BR (D) 3.1 and attachment) and an October 10, 1990 memo to Executive Deputy Commissioner Sheldon relating to "Authentic Assessment in Other States" were also consulted.

APPENDIX G

URBAN DISTRICT LEADERSHIP CONSORTIUM PROJECT PARTICIPATING DISTRICTS

Classroom-Centered

Performance-Based Assessment Project
in High School Science and Math

Participating Districts

Albuquerque, NM

Hammond, IN

Washington, DC

Dade County, FL

Kansas City, MO

Rochester, NY

Detroit, MI

San Francisco, CA

Chicago, IL

Newark, NJ

Los Angeles, CA

Cleveland, OH

New Orleans, LA

Cincinnati, OH

Pittsburgh, PA

Philadelphia, PA

St. Paul, MN

Baltimore, MD

District-Centered

Urban District Leadership Consortium and Philadelphia Schools Collaborative Alternative Assessments Project

Participating Districts

Albuquerque, NM

Boston, MA

Cincinnati, OH

Dade County, FL

Kansas City, MO

Los Angeles, CA

Newark, NJ

New Orleans, LA

New York, NY

Philadelphia, PA

Pittsburgh, PA

Rochester, NY

St. Paul, MN

San Francisco, CA

Source: Urban District Leadership Consortium. **Urban District Leadership Consortium Assessment Projects**. September 5, 1990.

BIBLIOGRAPHY

American Education Research Association, American Psychological Association, and National Council on Measurement in Education. **Standards For Educational and Psychological Testing**, Washington, D.C.: American Psychological Association, 1985.

Archbald, Doug and Fred M. Newmann. **Beyond Standardized Testing**, Reston, VA: National Association of Secondary School Principals, 1988.

Aschbacher, Pamela R. "CRESST Survey: State Activity and Interest in Performance-Based Assessment." Presented at 20th Annual Education Commission of the States/Colorado Department of Education Assessment Conference, Boulder, Colorado, June 1990.

Baron, Joan Boyoff. "Performance Testing In Connecticut." **Educational Leadership**, Vol. 47, No. 7, April 8, 1989, p. 8.

Blank, Rolf K. "State Policies and State Role in Curriculum," Washington, D.C.: Council of Chief State School Officers, pp. 20-22.

Burstall, Claire. "Innovative Forms of Assessment: A United Kingdom Perspective," **Educational Measurement: Issues and Practice**, Spring 1986, pp. 17-22.

"Business Leaders Urge Student Test of Workplace Skills." **Report on Education Research**, November 28, 1990, p. 4.

Cannell, John J. **Nationally Normed Elementary Achievement Testing in America's Public Schools: How All Fifty States Are Above the National Average**, Daniels, WV: Friends for Education, 1987.

Cannell, John J. **How Public Educators Cheat on Standardized Achievement Tests**, Albuquerque, NM: Friends for Education, 1989.

Coley, Richard and Margaret E. Goertz. **Educational Standards in the 50 States: 1990**, Princeton: Educational Testing Service, August 1990.

Corbett, H. Dickson and Bruce L. Wilson. **Unintended and Unwelcomed: The Local Impact of State Testing**, Philadelphia, PA: Research for Better Schools, May 1990.

Council of Chief State School Officers. "Accountability That Supports School Restructuring." **Success for All in a New Century**, November 1989, pp. 33-36, 39.

Council of Chief State School Officers. "Descriptions of Performance Assessment in the States," December 1990.

Council of Chief State School Officers. "Interstate Consortium on Alternative Methods of Student Assessment," December 1990.

Eckstein, Max A. and Harold J. Noah. "Forms and Functions of Secondary School-Leaving Examinations." **Comparative Education Review**, Vol. 33, No. 3, 1989, pp. 295-316.

Feinberg, Lawrence. "Multiple-Choice and Its Critics," **Commentaries**, The College Board, 1990.

"Grant Awarded for National Exam," **Education Week**, January 9, 1991, p. 14.

Gronlund, Norman E. **Measurement and Evaluation in Teaching**, New York, NY: 1985, pp. 13-16.

"Group To Help States Develop Alternate Tests," **Education Week**, December 5, 1990, p.4.

Haney, Walter and George Madaus. "Searching for Alternatives to Standardized Tests: Whys, Whats, and Whithers," **Phi Delta Kappan**, Vol. 70, No. 9, May 1989, pp. 683-687.

Harris, Joan and Joseph Sammons. **Failing Our Children: How Standardized Tests Damage New York's Youngest Children**. New York City: New York Public Interest Research Group, November 1989.

Martinez, Martin E. and Joseph I. Lipson. "Assessment for Learning." **Educational Leadership**, Vol. 46, No. 7, April 1989, pp. 73-75.

Medina, Noe and D. Monty Neill. **Fallout From the Testing Explosion**, Cambridge, MA.: National Center for Fair and Open Testing, June 1988.

Michigan Department of Education. **Association of State Assessment Programs Survey of Large-Scale Assessment Programs**, Lansing, Michigan: April 1990.

Mitchell, Ruth. "Performance Assessment: An Emphasis on 'Activity'," **Education Week**, January 24, 1990, pp. 36, 25.

Morganthau, Tom. "A Consumer's Guide to Testing." **Newsweek Special Edition**, Vol. CXVI. Fall/Winter 1990, pp. 62-63, 66, 68.

National Center for Fair and Open Testing. "Coalition Calls for 'Genuine Accountability'," **FairTest Examiner**, Winter 1989-90, pp. 1, 6, 7.

National Center for Fair and Open Testing. "New York Campaign Aims to End Testing in Grades P-2," **FairTest Examiner**, Summer 1990, pp. 8-10.

National Center for Fair and Open Testing. "Reformers Seek Halt to NAEP Expansion," **FairTest Examiner**, Summer 1990, pp. 7, 12.

National Center for Fair and Open Testing. "What is Authentic Evaluation?," **FairTest Examiner**, Winter 1990, pp. 8-9.

National Center for Fair and Open Testing. "Performance-Based Testing Spreads Across the Country." **FairTest Examiner**, Summer 1990, pp. 12-14, 16.

National Commission on Testing and Public Policy. **From Gatekeeper to Gateway: Transforming Testing in America,** Chestnut Hill, MA: National Commission on Testing and Public Policy, May 1990.

New York Public Interest Research Center and FairTest. **Standardized Tests and Our Children: A Guide to Testing Reform in New York.** New York City: New York Public Interest Research Center, 1990.

New York State Education Department. "A New Compact for Learning: Assessing Results," Report to the Board of Regents, October 1990.

New York State Education Department. "New York State Education Program for Elementary and Secondary Education," Report to the Board of Regents, March 1989.

New York State Education Department. "Other States and Authentic Assessment," Report to Executive Deputy Commissioner Thomas Sheldon, October 10, 1990.

Olson, Lynn. "MacArthur Awards $1.3 million for National Exams," **Education Week**, December 12, 1990, p. 5.

Postlethwaite, T. Neville. "The Use of Standardized Tests in Secondary Schools in Four European Countries," **ERIC**, ED 384 455, November 1986.

Rothman, Robert. "Commercial Test Publishers Entering Market for Performance Assessments," **Education Week**, September 12, 1990, p. 11.

Rothman, Robert. "New Tests Based on Performance Raise Questions." **Education Week**, September 12, 1990, pp. 1, 10, 12.

Rothman, Robert. "U.C.L.A. Gets $14.3 Million for Research Center on Testing," **Education Week**, November 21, 1990, p. 25.

Rothman, Robert. "U.S. Opts Out of International Performance Assessment," **Education Week**, September 12, 1990, p. 10.

Shepard, Lorrie A. "Why We Need Better Assessments," **Educational Leadership**, Vol. 46, No.7, April 1989, pp. 4-9.

Spofford, Tim. "Changes in State's Testing of Pupils Urged." Albany **Times Union**, October 18, 1990, p. B-11.

"Standardized Tests Will Bow to Tests of Outcomes, ETS Says," **Report of Education Research**, November 28, 1990, pp. 3-4.

Urban District Leadership Consortium. **Urban District Leadership Consortium Assessment Projects**, Washington, D.C.: American Federation of Teachers Center for Restructuring, September 5, 1990.

U.S. Congress Office of Technology Assessment. **Schools, Kids, and Measurement: Technologies of Assessment** (project description), Washington, D.C.: Office of Technology Assessment, 1990.

U.S. Department of Education. Office of Educational Research and Improvement. **Creating Responsible and Responsive Accountability Systems**, Washington, D.C.: U.S. Department of Education, September 1988, pp.40-41.

van den Bergh, Huub. "Large Scale Oracy Assessment in the Netherlands," **ERIC**, ED 292 834, 1987.

Voss, Bernd. "Language Testing in German Secondary Schools--Theory and Practice," **ERIC**, ED 266 691, 1985.

Wiggins, Grant. "A True Test: Toward More Authentic and Equitable Assessment," **Phi Delta Kappan**, Vol. 70, No. 9, May 1989.

Wiggins, Grant. "Teaching to the (Authentic) Tests." **Educational Leadership**, Vol. 46, No. 7, April 1989, pp. 41-47.

OTHER INFORMATION SOURCES

TELEPHONE INTERVIEWS CONDUCTED:

Armstrong, Jane. Director of Policy Studies, Education Commission of the States, Denver, Colorado (November 14, 1990).

Aschbacher, Pam. Project Director, Center for Research, Evaluation, Standards and Student Testing, University of California at Los Angeles (October 11, November 14, December 7 and December 13, 1990).

Blank, Rolf. Associate Director of Assessment, Council of Chief State School Officers, Washington, D.C. (October 12, 1990).

Bracken, Morgan. Staff, International Assessment of Educational Progress (IAEP), Educational Testing Service, Princeton, New Jersey (October 16, 1990).

Bush, Melody. Information Specialist, Education Commission of the States, Denver, Colorado (December 4, 1990).

Cawthorne, John. UDLC Project Director, Center for the Study of Testing, Evaluation and Educational Policy, Boston College, Chestnut Hill, Massachusetts (December 12, 1990).

Frelow, Fred. Research Associate, Center for School Reform, Teachers College, Columbia University, New York, New York (October 16, 1990).

Fulton, Kathleen. Analyst, Office of Technology Assessment, U.S. Congress (October 26 and December 7, 1990).

Goldberg, Bruce. Coordinator, Urban District Leadership Consortium, American Federation of Teachers' Center for Restructuring, Washington, D.C. (October 8, 1990).

Neill, Monty. Associate Director, National Center for Fair and Open Testing (FairTest), Cambridge, Massachusetts (October 16, October 22 and December 12, 1990).

Mead, Nancy. Director, International Assessment of Educational Progress (IAEP), Educational Testing Service, Princeton, New Jersey (October 17, 1990).

Mitchell, Ruth. Associate Director, Council for Basic Education, Washington, D.C. (October 17, 1990).

Pipho, Chris. Director, Clearinghouse/State Relations, Education Commission of the States, Denver, Colorado (October 16 and November 20, 1990).

Power, Nancy. Staff, Nelson Canada (Test Publisher) Measurement and Guidance Division, Scarborough, Ontario, Canada (December 4, 1990).

Purves, Alan. Professor, School of Education, State University of New York at Albany (October 16, 1990).

Wilson, Bruce L.. Staff, Research for Better Schools, Philadelphia, Pennsylvania (October 8, 1990).

CORRESPONDENCE RECEIVED:

Beckelhimer, Steve, Assistant Director, Student Assessment Services, West Virginia Department of Education (November 15, 1990).

Cusack, Jacqueline L., Acting Assistant Commissioner, Division of General Academic Education, State of New Jersey Department of Education (November 27, 1990).

Gabbert, Larry, State Supervisor Educational Assessment, Research and Evaluation Division, State of Delaware Department of Public Instruction (November 9, 1990).

Pernici, Sam C., Education Bureau Assistant Administrator, Bureau of Pupil Accountability, State of Louisiana Department of Education (November 9, 1990).

Power, Nancy, Product Manager, Measurement and Guidance Division, Nelson Canada, Scarborough, Ontario, Canada (December 14, 1990).

Trimble, C. Scott, Director, Division of Evaluation, Kentucky Department of Education (December 4, 1990).

91-6990